

**Eye Tracking:
Empirische Ableitung und quantitative Analyse
eines Indikators für Target Fixations
im Hubschraubersimulator**

Dissertation
zur Erlangung des akademischen Grades eines
Doktors der Sicherheitswissenschaften (Dr. rer. sec.)

im
Fachbereich D – Architektur, Bauingenieurwesen,
Maschinenbau, Sicherheitstechnik

der
Bergischen Universität Wuppertal
- Abteilung Sicherheitstechnik -



vorgelegt von
Dipl.-Psych. Maxi Robinski
aus Chemnitz

Wuppertal 2013

Die Dissertation kann wie folgt zitiert werden:

urn:nbn:de:hbz:468-20140508-125442-3

[<http://nbn-resolving.de/urn/resolver.pl?urn=urn%3Anbn%3Ade%3Ahbz%3A468-20140508-125442-3>]

„Wer weiß, ob die Gedanken nicht auch einen ganz winzigen Lärm machen, der durch feinste Instrumente aufzufangen und empirisch (durch Vergleich und Experiment) zu enträtseln wäre.“

(Morgenstern, 1896)

Der größte Dank gilt Ihnen, PD Dr. Michael Stein, für die Betreuung der Dissertation, die fachliche und ideelle Unterstützung sowie Ihr Vertrauen. Nicht nur, dass Sie mich wegweisend in das Gebiet der Ergonomie eingearbeitet haben, Sie ermöglichten mir auch tiefgründigen Diskurs weit über die Grenzen dieser Arbeit hinaus.

Aufrichtig danke ich auch der Leitung des Flugmedizinischen Instituts der Luftwaffe für den Forschungsfreiraum und die Ermöglichung dieses Promotionsvorhabens.

Nachhaltig dankbar bin ich der Leitung und dem Fliegerarzt des Simulatorzentrums der Heeresfliegerwaffenschule Bückeburg für die Kooperation und Unterstützung der Studie.

Kaum dankbarer könnte ich den Teilnehmern der Untersuchung sein, Fluglehrer der Heeresfliegerwaffenschule und Lehrgruppe A, III. Inspektion, die sich als auskunftsbereite Experten und geduldige Probanden bewiesen.

Eure Unterstützung war es, liebe Kollegen der Abteilung Ergonomie des Flugmedizinischen Instituts der Luftwaffe, die mich zum konzentrierten Arbeiten befähigte. Ich danke Andreas Herma für seine umsichtige, technische Assistenz sowie den begnadeten Korrekturleserinnen Katja-Bianka Catizone, Eleonore Neukäufer, Dagmar Krestel und Ramona Städtler. Auch danke ich Karl-Heinz Hintzen für seine engagierten Antworten aus der Perspektive des Ingenieurs.

Für meine Großeltern Dr. Renate Robinski und Dr. habil. Claus Robinski[†].

Zusammenfassung

Einleitung: Die Ausbildungskosten für einen Hubschrauberpiloten betragen etwa EUR 2.3 Mio. Entsprechend sollte die Ausbildung so effizient wie möglich realisiert werden. Studien aus der internationalen Militärluftfahrt zeigen, dass Eye Tracking als Feedbacktool für Piloten den Ausbildungsfortschritt bereichern kann, indem Scanning Techniken als Indikatoren der kognitiven Informationsverarbeitung von erfahrenen Piloten analysiert, visualisiert und an unerfahrene Piloten vermittelt werden können. Für den Luftfahrtbereich liegen bislang wenige wissenschaftliche Erkenntnisse zu dieser Methode vor. Die Dissertation beschäftigt sich daher mit dem Potential der Eye Tracking Methode zur Analyse und Aufbereitung der Scanning Techniken von Hubschrauberpiloten.

Methoden: Nach Reflektion der technischen, theoretischen und empirischen Hintergründe im Bereich Eye Tracking wurde ein mehrfaktorielles Untersuchungsdesign angelegt, innerhalb dessen der Einfluss von Flugerfahrung und visuellen Anforderungen einer Flugaufgabe auf Scanning Techniken von Piloten geprüft wurde. Nach Durchführung von A-Priori-Experteninterviews ($N = 15$) zu den Inhalten der Flugaufgaben und einem Machbarkeitsversuch nahmen an der Hauptuntersuchung 33 Hubschrauberpiloten ($n_{\text{Fluglehrer}} = 17$, $n_{\text{Flugschüler}} = 16$) teil, die, ausgestattet mit dem Eye Tracking System Dikablis der Firma Ergoneers GmbH, jeweils zwei verschiedene Missionen in einem Hubschraubersimulator absolvierten. Neben der objektiven Messung von Blickparametern wurden auch die Workload sowie die subjektive Performance der Luftfahrzeugführer und eine Selbsteinschätzung zu deren Scanning Techniken erfasst. Darüber hinaus bearbeiteten alle Probanden einen Fragebogen zur ergonomischen Bewertung der Eye Tracking Methode als Feedbacktool. Bei der Aufbereitung der Daten wurde zur Operationalisierung von Scanning Techniken ein Indikator für Blickfixierungen (Target Fixations) gebildet, welcher für Blicke auf die Cockpitinstrumente auf Empfehlungen von Luftfahrtbehörden basierte und für Blicke in die Außenwelt statistisch aus der Stichprobe errechnet wurde.

Ergebnisse: Die Ergebnisse zeigen, dass sich Scanning-bezogene Unterschiede zwischen Hubschrauberpiloten zu 54 % auf die Faktorkombination aus Expertise und visuellen Anforderungen zurückführen lassen ($F = 4.88$, $p = .033$; Target Fixations: $\eta^2 = .12$; mittelstarker Effekt). Bei der Gegenüberstellung stellte sich heraus, dass erfahrene Piloten ihr peripheres Sichtfeld je nach visuellen Anforderungen effektiver nutzten und je nach Flugphase andere Informationskanäle als Flugschüler zur Flugführung einsetzten. Beim Vergleich objektiver Eye Tracking Daten und subjektiver Einschätzungen fiel auf, dass sich nicht nur unerfahrene Piloten zum Teil verschätzten und ihnen ihr tatsächliches Scanning Verhalten nicht immer eindeutig bewusst war. Die Piloten überschätzten beispielsweise den Anteil ihrer Blicke nach außen (Flugschüler um 26 %, Fluglehrer um 10 %). In diesem Zusammenhang wurde eine statistische Beziehung zwischen der Korrektheit der Selbsteinschätzung der Scanning Techniken und der Performance beim Durchführen von Landemanövern gefunden ($r = .31$, $p = .035$). Die Piloten bewerteten die Eye Tracking Methode zu 83 % als nicht hinderlich für eine sichere Flugführung und konnten sich zu knapp zwei Dritteln vorstellen, das Verfahren regelmäßig einmal im Monat bei der Ausbildung als Feedbacktool einzusetzen.

Diskussion: Die Ergebnisse implizieren, dass Eye Tracking den Kriterien einer wissenschaftlichen Methode zur Analyse der visuellen Informationsakquise von Hubschrauberpiloten weitestgehend gerecht wird: Zum einen lassen sich signifikante Unterschiede zwischen unterschiedlich erfahrenen Piloten abbilden und zum anderen lässt sich ein detailliertes Bild über den strategischen Einsatz der Scanning Techniken von Experten zeichnen. Somit eröffnet sich eine neue Möglichkeit dafür, das Ausbildungs- und Trainingssystem für Piloten aber auch ergonomische Forschungsanliegen innerhalb der Luftfahrt methodisch zu untermauern. Dennoch besteht hinsichtlich der technischen Systemkomponenten dahingehend Entwicklungsbedarf, dass durch Folgestudien der Einsatz von Eye Tracking im Realflug geprüft werden sollte. Entsprechende Ansatzpunkte werden aufgegriffen und diskutiert.

Summary

Introduction: The training costs for a single helicopter pilot amount to approximately 2.3 million EUR. It is thus preferable to implement training as efficiently as possible. Studies in international military aviation show that eye tracking as a feedback tool for pilots can enhance the progress of training when analyze and visualize scanning techniques of experienced pilots as indicators of cognitive information processing and passing them on to inexperienced pilots. However, to date scientific findings on this method are rare with regard to aviation. Therefore, the dissertation addresses the potential of the eye tracking method with respect to the analysis and processing of helicopter pilot scanning techniques.

Methods: After a critical reflection on the technical, theoretical and empirical background of eye tracking a multi-factor experimental design was established to verify the impact of flight experience and the visual demands of a flight task on pilot scanning techniques. Following a priori expert interviews ($N = 15$) on the contents of the flight tasks and a feasibility test 33 helicopter pilots participated in the main experiment ($n_{\text{instructor pilots}} = 17$, $n_{\text{student pilots}} = 16$) and performed two different missions in a helicopter simulator using the Dikablis eye tracking system by Ergoneers GmbH. In addition to the objective measurement of gaze parameters workload as well as subjective pilot performance and a self-assessment of their scanning techniques were considered in the investigation. Furthermore, all subjects completed a questionnaire on the ergonomic evaluation of the eye tracking method as a feedback tool. To operationalize scanning techniques, an indicator for target fixations was generated during the processing of eye tracking data; this indicator was based on recommendations by aviation authorities with respect to gazes at the flight instruments and statistically calculated on the basis of the sample with regard to gazes at the outside world.

Results: The results show that 54 % of scanning-related variance between helicopter pilots can be attributed to the factor combination of expertise and visual demands ($F = 4.88$, $p = .033$; target fixations: $\eta^2 = .12$; medium effect size). A comparison revealed that experienced pilots utilize their peripheral field of view more effectively depending on visual demands and strategically use other information channels (flight instruments vs. external reference points) than student pilots depending on the individual flight phases. The comparison of objective eye tracking data and subjective assessments showed that judgment errors are not only made by inexperienced pilots and that they are not always clearly aware of their actual scanning techniques. For example, the proportion of scans in the outside world was overestimated on average (student pilots around 26 %, instructor pilots around 10 %). In this context, a statistical relation between the correctness of the self-assessment of scanning techniques and the pilots' performance when conducting landing maneuvers was found ($r = .31$, $p = .035$). 83 % of the pilots did not regard the eye tracking method as a limitation to flight safety and nearly two thirds could imagine using this method as a feedback tool for training once a month on a regular basis.

Discussion: The results imply that eye tracking largely fulfills the criteria of a scientific method for the analysis of visual information acquisition by pilots: On the one hand, significant differences between pilots with varying degrees of experience can be described, and on the other hand a detailed picture of the strategical use of scanning techniques by experts can be provided. This offers a new option of methodically substantiating the pilot training system and also ergonomic research interests. Nevertheless, there is a development requirement for technical system components to the effect that the use of eye tracking in real flights should be verified in follow-up studies. Appropriate approaches are being addressed and discussed.

Inhaltsverzeichnis

1	EINLEITUNG	1
1.1	DIE AUSBILDUNG VON HUBSCHRAUBERPILOTEN.....	1
1.2	EYE TRACKING IN DER LUFTFAHRT – QUO VADIS?.....	3
1.3	IMPETUS DER ARBEIT	6
2	TECHNISCHE, THEORETISCHE UND EMPIRISCHE HINTERGRÜNDE.....	9
2.1	TECHNISCHE GRUNDLAGEN ZU EYE TRACKING	9
2.1.1	Funktionsweise von Eye Tracking Systemen.....	9
2.1.2	Kalibrierung und Messgenauigkeit von Eye Tracking Systemen	11
2.1.3	Arten von Eye Tracking Systemen.....	14
2.1.4	Heat Maps.....	16
2.1.5	Überblick der Nach- und Vorteile von Eye Tracking	17
2.1.6	Blickparameter und ihre Interpretation	19
2.1.7	Abgrenzung zu anderen okulografischen Messmethoden	22
2.1.8	Anwendungsbereiche von Eye Tracking	25
2.1.9	Zusammenfassung technischer Aspekte.....	27
2.2	THEORETISCHE UND EMPIRISCHE GRUNDLAGEN ZU EYE TRACKING	27
2.2.1	Flugführung als anspruchsvolle Prozessführung.....	27
2.2.2	Ein Grundmodell der Informationsverarbeitung	28
2.2.3	Ergänzungen zum Modell des „Information Processors“	31
2.2.4	Einfluss affektiver und motivationaler Komponenten	32
2.2.5	Aufmerksamkeit und Wahrnehmung	35
2.2.6	Visuelle Aufmerksamkeit	41
2.2.7	Visuelle Aufmerksamkeit in der Luftfahrt.....	45
2.2.8	Arbeitsgedächtnis, Belastung, Beanspruchung und Workload.....	52
2.2.9	Aufgabenanforderungen: Komplexität vs. Schwierigkeit einer Mission.....	62
2.2.10	Erfahrung und Langzeitgedächtnis	67
2.2.11	Situationsbewusstsein	71
2.2.12	Schlussfolgerungen für die Untersuchung.....	75
2.3	FORSCHUNGSFRAGEN UND HYPOTHESEN	79
2.3.1	Multivariates Modell.....	79
2.3.2	Explorative Fragestellungen.....	80
3	METHODEN.....	81
3.1	UNTERSUCHUNGSABLAUF UND EXPERIMENTALPLAN	81
3.1.1	Machbarkeitsversuch.....	82
3.1.2	Hauptversuch.....	83
3.2	EQUIPMENT.....	85
3.2.1	Simulator EC 135.....	85
3.2.2	Eye Tracking System	87
3.3	OPERATIONALISIERUNG DER KONSTRUKTE UND ZUGEHÖRIGE UNTERSUCHUNGSINSTRUMENTE	93
3.3.1	Operationalisierung der unabhängigen Variablen.....	93
3.3.2	Operationalisierung der abhängigen Variablen	97
3.4	DATENAUSWERTUNG.....	101
3.4.1	Hypothesengeleitete Datenauswertung.....	101
3.4.2	Explorative Datenauswertung	102
3.5	STICHPROBE.....	103

3.5.1	„Optimale“ Stichprobe.....	103
3.5.2	Tatsächliche Stichprobe.....	103
4	ERGEBNISSE	105
4.1	ERGEBNISSE DES MACHBARKEITSVERSUCHS.....	105
4.1.1	Implementierbarkeit des Eye Tracking Systems	105
4.1.2	Validität der Blickdaten.....	106
4.1.3	Qualitative Auswertungsbeispiele für Einzelfallanalysen	107
4.1.4	Zusammenfassung des Machbarkeitsversuchs.....	108
4.2	ERGEBNISSE DES HAUPTVERSUCHS	108
4.2.1	Deskriptive Auswertung.....	108
4.2.2	Inferenzstatistische Auswertung	112
4.2.3	Explorative Auswertung.....	118
4.2.4	Zusammenfassung des Hauptversuchs.....	128
5	DISKUSSION	129
5.1	SCANNING TECHNIKEN VON HUBSCHRAUBERPILOTEN.....	129
5.2	NÜTZLICHKEIT DER EYE TRACKING METHODE	131
5.3	WISSENSCHAFTLICHKEIT DER EYE TRACKING METHODE	133
5.4	GENERALISIERBARKEIT UND EINSCHRÄNKUNGEN DER STUDIE.....	136
5.4.1	Zur Rolle von Target Fixations nach Rücksprache mit Experten	136
5.4.2	Zur Rolle von Performance und Workload	138
5.4.3	Zum Versuchsaufbau	141
5.4.4	Zum Eye Tracking System Dikablis Cable	143
5.5	PRAKTISCHE IMPLIKATIONEN UND FORSCHUNGS-AUSBLICK.....	144
5.5.1	Training	144
5.5.2	Auswahl.....	145
5.5.3	Forschung.....	145
6	LITERATURVERZEICHNIS.....	147
7	ANHANG	159
7.1	ABKÜRZUNGSVERZEICHNIS.....	159
7.2	EXPERTENINTERVIEW	161
7.3	RANGLISTE.....	163
7.4	GENEHMIGUNGSSCHREIBEN BMVG	165
7.5	PROBANDENINFORMATION	167
7.6	INTERVIEW FÜR FLUGLEHRER.....	169
7.7	INTERVIEW FÜR FLUGSCHÜLER.....	171
7.8	FRAGEBOGEN ZUR ERGONOMISCHEN BEWERTUNG DER EYE TRACKING METHODE	173
7.9	VALIDITÄTSINDIZES FÜR PUPILLE UND MARKER	177
7.10	ERGEBNISSE ZUM NASA-TLX.....	179
7.11	POST HOC ANALYSEN ZUR MANOVA	180
8	LEBENS-LAUF DER AUTORIN.....	183

Tabellenverzeichnis

Tabelle 1: Untergrenzen für die Fixationsdauer.	49
Tabelle 2: Obergrenzen für die Fixationsdauer (nur im Fahrkontext).	51
Tabelle 3: Experimentalplan.	82
Tabelle 4: Technische Daten zum Flugsimulator EC 135.....	87
Tabelle 5: Software-Komponenten.....	89
Tabelle 6: Rangliste der fünf Landemanöver (N = 15).	96
Tabelle 7: Cutoff-Werte für Target Fixations.	98
Tabelle 8: Scanning Techniken (gemittelt) in AOIs.	106
Tabelle 9: Mittlere Blickhäufigkeiten und Blickdauer für OTW und Instrumente.....	109
Tabelle 10: Deskriptive Kennwerte für Target Fixations.	110
Tabelle 11: Deskriptive Kennwerte für Performance.....	110
Tabelle 12: Deskriptive Kennwerte für Workload (NASA-TLX [%]).	111
Tabelle 13: Überprüfung der statistischen Voraussetzungen.	112
Tabelle 14: MANOVA zur Hypothesenprüfung.	113
Tabelle 15: Aufklärungsgüte der MANOVA bei Variation des Cutoff-Wertes.	115
Tabelle 16: MANOVA für verschiedene Flugphasen.	117
Tabelle 17: Korrelationen zwischen Performance und Blickparametern (N = 45).	121
Tabelle 18: Abschlussfragebogen (N = 29).....	125
Tabelle 19: Nützlichkeitsbewertung.....	127
Tabelle 20: Zusammenfassende Bewertung.....	135
Tabelle 21: Leitfaden zum standardisierten Vorgehen bei Eye Tracking Studien.	146
Tabelle 22: Validitätsindizes für Pupillen- und Markervalidität (N = 33).	177
Tabelle 23: Subskalen des NASA-TLX (N = 66).	179
Tabelle 24: T-Test für die Subskalen des NASA-TLX auf dem Faktor Erfahrung.	179
Tabelle 25: Target Fixations auf dem Faktor visuelle Anforderungen.	181
Tabelle 26: Follow Up ANOVA auf dem Faktor visuelle Anforderungen.....	181
Tabelle 27: Target Fixations auf dem Faktor Erfahrung.....	181
Tabelle 28: Follow Up ANOVA auf dem Faktor Erfahrung.	181

Abbildungsverzeichnis

Abbildung 1: Simulator und Cockpit Mock-Up des Eurocopter (EC) 135 (eigene Fotografien).	1
Abbildung 2: Anflug und Landung im Gelände (übernommen aus Grube, 2007).	2
Abbildung 3: Simulierte Landungen von unerfahrenen (links) und erfahrenen (rechts) Zivilpiloten (übernommen aus Kasarskis et al., 2001).	3
Abbildung 4: Integriertes Helmsichtsystem.....	4
Abbildung 5: Prinzip der cornealen Reflexion und Purkinje-Reflexe9	9
Abbildung 6: Projektion des Blickpunktes auf die Blickebene ($X, Y \in \mathbb{R}^2$; $X_L Y$; eigene Fotografie).	10
Abbildung 7: Kalibrierradius beim Eye Tracking System Dikablis.....	11
Abbildung 8: Trigonometrisches Berechnungsprinzip der Messgenauigkeit im Flugsimulator.	12
Abbildung 9: Versuchsaufbau bei Pfeiffer et al. (2007; übernommen aus ebd.).....	13
Abbildung 10: Head Mounted System.	14
Abbildung 11: Remote System (übernommen aus DeSantis et al., 2005).	14
Abbildung 12: Heat Map.....	16
Abbildung 13: Grundmodell der Informationsverarbeitung (übersetzt aus Wickens, 2009).	29
Abbildung 14: Navigation im Modell von Wickens (2009): Aufmerksamkeit und Wahrnehmung.	35
Abbildung 15: Flaschenhalstheorie.....	36
Abbildung 16: 4D-Modell multipler Ressourcen (übernommen aus Wickens, 2008).	38
Abbildung 17: Scanning Techniken im Modell von Wickens (2009).	41
Abbildung 18: Klassifikation der Blickdauer nach Zwahlen et al. (1988).	50
Abbildung 19: Navigation im Modell von Wickens (2009): Arbeitsgedächtnis.	52

Abbildung 20: Zusammenhang Belastung und Beanspruchung (modifiziert aus Grandt, 2004).	53
Abbildung 21: Verfügbare vs. benötigte Ressourcen (übernommen aus Flemisch, 2001).	60
Abbildung 22: Navigation im Modell von Wickens (2009):.....	68
Abbildung 23: Navigation im Modell von Wickens (2009): Situationsbewusstsein.	72
Abbildung 24: Modell des Situationsbewusstseins (übernommen aus Endsley, 1995).	72
Abbildung 25: Arbeitsmodell für die Eye Tracking Studie.....	79
Abbildung 26: Untersuchungsablauf.	81
Abbildung 27: Fixierung der Komponenten des Eye Tracking Systems (eigene Fotografie).	84
Abbildung 28: EC 135 Hubschrauber und Simulatordom (eigene Fotografie).	86
Abbildung 29: Externe Konsole (eigene Fotografie).	86
Abbildung 30: Head Unit.....	87
Abbildung 31: Elektronikeinheit und Infrarot-Marker (eigene Fotografien).	88
Abbildung 32: Skizzierter Versuchsaufbau (gestrichelte Linien: Kabelverbindungen).	89
Abbildung 33: Aufnahmesoftware während Pupillenerkennung (aus: Dikablis Recording).	90
Abbildung 34: Kalibrierpunkte (eigene Fotografie).	90
Abbildung 35: Dikablis Player.	91
Abbildung 36: Dikablis Analysis während der Pupillenerkennung.....	91
Abbildung 37: D-Lab Auswertungsoberfläche.	93
Abbildung 38: Missionen (links: Pinnacle, rechts: Gruppenversorger, aus: Dikablis Player).	97
Abbildung 39: Platzrunde und Landepunkte (nicht maßstabgetreu).....	97
Abbildung 40: Ex-ante Bestimmung der Stichprobengröße.	103
Abbildung 41: Flugerfahrung (Unterschied zwischen den Gruppen: $t[28] = 7.12, p < .000$).....	104
Abbildung 42: Simulatorerfahrung (Unterschied zwischen den Gruppen: $t[28] = 4.80, p < .000$).....	104
Abbildung 43: Proband mit Brille (eigene Fotografien).	105
Abbildung 44: Versuchsleiterplatz und Verkabelung der Controlling Box (eigene Fotografien).....	105
Abbildung 45: Verrutschen der Brille (eigene Fotografien).	106
Abbildung 46: Proband 1 vs. 2 während Landung auf dem Gruppenversorger (aus: D-Lab).....	107
Abbildung 47: Heat Map für Proband 1 für Mission 1 vs. 2 (aus: D-Lab).	107
Abbildung 48: Univariate Normalverteilung der AVs (Target Fixations, Performance, Workload)....	112
Abbildung 49: Interaktionseffekt der Faktoren für die Variable Target Fixations.	114
Abbildung 50: Haupteffekt des Faktors Erfahrung für die Variable Performance.	114
Abbildung 51: Relativer Anteil an allen Target Fixations über AOIs und Flugphasen für Pinnacle. ...	118
Abbildung 52: Relativer Anteil an allen Target Fixations über AOIs und Flugphasen für Gruppenversorger.	118
Abbildung 53: Differenzvariablen (links: OTW, rechts: Instrumente; N = 45).....	119
Abbildung 54: Fehleinschätzungen für OTW in Abhängigkeit der Erfahrung.....	120
Abbildung 55: Fehleinschätzungen für Instrumente in Abhängigkeit der Erfahrung.	121
Abbildung 56: Regression Performance auf Selbsteinschätzung beim Instrumenten Check.....	122
Abbildung 57: Selbsteinschätzung besserer (n = 9) und schlechterer Performer (n = 11).....	123
Abbildung 58: Druckstellen nach Tragen der Brille (eigene Fotografien).	124
Abbildung 59: Eye Tracking als Feedback-Methode (N = 29).....	126
Abbildung 60: Eignung von Eye Tracking für den Realflug.....	127
Abbildung 61: Cockpit Navy H-60 mit LED-Unterstützung.	138
Abbildung 62: Subskalen des NASA-TLX (N = 66).....	179
Abbildung 63: Unterschied für die Subskala mentale Anforderungen auf dem Faktor Erfahrung. ...	180
Abbildung 64: Target Fixations auf dem Faktor visuelle Anforderungen.	180

1 Einleitung

1.1 Die Ausbildung von Hubschrauberpiloten

Die Hubschrauberführung ist eine komplexe Aufgabe und bewegt sich im Spannungsfeld zwischen sicherer Steuerung des nicht eigenstabilen Luftfahrzeugs sowie effizienter Realisierung eines Einsatzauftrags. Sie ist nicht nur innerhalb der deutschen fliegenden Streitkräfte mit umfangreichem Ausbildungsaufwand verbunden. Die Kosten für die Ausbildung eines Hubschrauberpiloten¹ betragen insgesamt etwa EUR 2.3 Mio., wobei etwa 50 % der 140 Flugstunden im Simulator (s. Abbildung 1) realisiert werden, da eine Simulatorstunde nur ein Viertel einer Realflugstunde kostet². Der dennoch hohe Gesamtkostenaufwand verlangt eine effiziente Realisierung der Ausbildung, bei der Pilotenanwärter so schnell wie möglich in die Lage versetzt werden sollen, wie Experten der Hubschrauberführung zu denken und zu handeln. Während des Trainings stehen Fluglehrer dabei häufig vor der Frage, wann und wie sie ihren Schülern die geeignetste Rückmeldung über ihre Leistung geben können. Die Situation im Flugsimulator, bei der der Fluglehrer neben dem Flugschüler im Cockpit sitzt und ihm die Flugführung vermittelt, bietet nur bedingt Möglichkeiten, die Handlungen des Flugschülers detailliert zu überwachen und ihm seine Fehler zu erklären (Sullivan, Yang, Day, & Kennedy, 2011).



Abbildung 1: Simulator und Cockpit Mock-Up des Eurocopter (EC) 135 (eigene Fotografien).

Beim Führen eines Luftfahrzeugs spielen adäquate Informationsverarbeitungsprozesse, die durch die Aufmerksamkeitsallokation, die Erfahrung von Piloten und die Anforderungen durch die Flugaufgabe moduliert werden, eine entscheidende

¹ Die weibliche Form ist der männlichen in dieser Arbeit gleichgestellt; lediglich aus Gründen der Übersichtlichkeit wurde die männliche Form gewählt.

² Oberstleutnant Pfähler, Heeresfliegerwaffenschule Bückeburg; Email-Kommunikation am 21.09.2012.

Rolle. Moderne Luftfahrzeuge beinhalten in den neuen „Glass Cockpits“ teilweise hochkomplexe Informations- und Anzeigesysteme, sodass die Informationsverarbeitung durch insbesondere kognitiven Ressourceneinsatz und vermehrte Blicke ins Cockpit³ gekennzeichnet ist (Colvin, Dodhia, & Dismukes, 2005). Entsprechend erhöht sich die Notwendigkeit des Trainings von effektiven Scanning Techniken zur Informationsakquise, denn visuelle Aufmerksamkeit ist die wichtigste Ressource von Piloten (European Aviation Safety Agency, EASA, 2010). Jones und Endsley (1996) konnten beispielsweise zeigen, dass 76 % der Mängel im Situationsbewusstsein von Piloten auf die unzureichende visuelle Wahrnehmung von relevanten Informationen zurückzuführen sind. Bisher ist dabei noch nicht konsistent geklärt, welche Scanning Techniken es sind, die Experten von Novizen unterscheiden (Sullivan et al., 2011). Am Beispiel der hohen Anforderungen eines Landemanövers im Gelände (s. Abbildung 2) wird die Notwendigkeit von Expertise und Übung deutlich, da bei der Entscheidung zwischen Absetzen oder Durchstarten wegen eines ungünstigen Landewinkels (Wind, unklare Bodenverhältnisse usw.) geschulte visuelle Informationsaufnahme und -verarbeitung obligat ist.

Anflug und Landung auf Damm und Kuppe

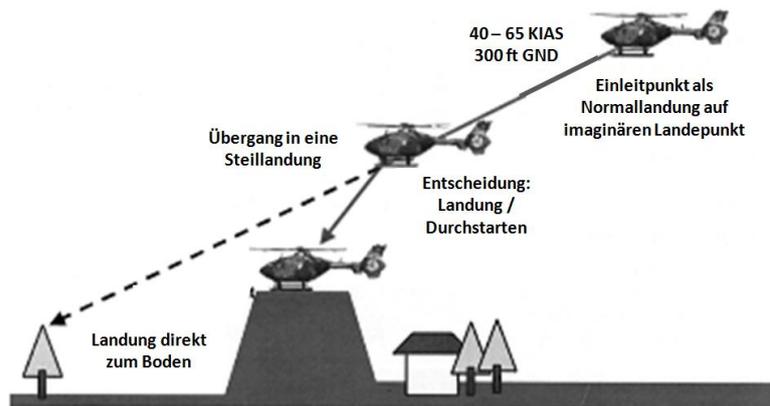


Abbildung 2: Anflug und Landung im Gelände (übernommen aus Grube, 2007).

Kasarskis, Stehwin, Hickox, Aretz und Wickens (2001) demonstrieren den Leistungsunterschied von erfahrenen und unerfahrenen Zivilpiloten bei einem simulierten Landemanöver grafisch (s. Abbildung 3). Sie zeigen, dass der Schwerpunkt der Verteilung der Landepunkte von Experten näher am optimalen Landepunkt (im Koordinatenursprung) liegt und deren Leistung damit im Mittel eine geringere Streuung aufweist als die von Novizen.

³ Blicke ins Cockpit werden auch als Head Down Anteile bezeichnet. Analog heißen Blicke in den Luftraum auch Head Up Anteile oder werden mit OTW (Out The Window) abgekürzt.

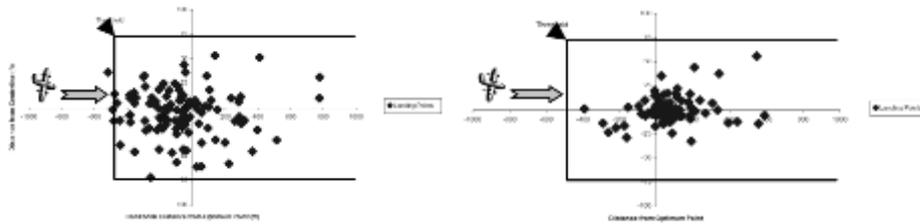


Abbildung 3: Simulierte Landungen von unerfahrenen (links) und erfahrenen (rechts) Zivildiplomaten (übernommen aus Kasarskis et al., 2001).

Studien aus der militärischen Luftfahrt (USA) zeigen, dass es Eye Tracking ermöglicht, die Leistung im Simulator bei Kampfpiloten (F-16 Flugschüler; Wetzel, Anderson, & Barelka, 1998) aber auch bei Hubschrauberpiloten (Navy H-60; Sullivan et al., 2011) zu steigern, weil Scanning Techniken von Flugschülern bei unterschiedlich komplexen Manövern detailliert aufgezeichnet und mit denen von Experten verglichen werden können. Hier bieten sich im Gegensatz zu reinen Videoanalysen Möglichkeiten, Scanning Fehler oder ausgelassene, wichtige Blicke von Flugschülern zu identifizieren und ihnen über individuelle visuelle Rückmeldung eine korrekte Scanning Technik zu vermitteln: „[...] *performance in the visual periphery can be improved through training by factors of as much as three [...]*“ (Strasburger, Rentschler, & Jüttner, 2011, S. 44). Dieses Prinzip wurde durch Shapiro und Raymond (1989) im Bereich Videospiele bereits statistisch demonstriert: Die Autoren fanden eine signifikante Korrelation zwischen Scanning Techniken und der Leistung beim Videospiele. Zwei Gruppen von Probanden lernten effizient bzw. ineffizient zu scannen. Effizientes Scanning bedeutete hier, dass Blickbewegungen minimiert, aber Blickwege optimiert werden sollten. Die ineffiziente Gruppe mit überflüssigen Blickbewegungen und falschen Blickwegen hatte im Ergebnis die gleiche Leistung wie eine Kontrollgruppe ohne Training.

1.2 Eye Tracking in der Luftfahrt – Quo vadis?

Damit „richtige“ Scanning Techniken an Arbeitsplätzen von Piloten effektiv und sicherheitsförderlich trainiert werden können, ist es nötig, dass Luftfahrzeugführer diese zunächst reflektieren und bewusst steuern können. Das Prinzip der bewussten Steuerung automatisierter, physiologischer Reaktionen ist aus dem Bereich Biofeedback hinlänglich bekannt: Erhalten Personen visuelle Rückmeldung über biologische Reaktionen (hier zum Beispiel über Scanning Techniken), erhöht sich das Bewusstsein hierüber, steigt die Motivation zur Veränderung bzw. Verbesserung der Reaktionen und entsteht operante Verhaltenskontrolle (z. B. Janelle & Hatfield, 2008). Innerhalb der Luftfahrt gehört ein visuell unterstütztes Feedbacktraining zu Scanning Techniken bislang noch nicht explizit zum Trainingsplan von Piloten. Hier zeichnen sich Ansatzpunkte für Weiterentwicklungen ab.

Neben dem bisherigen Fehlen von wissenschaftlichen Erkenntnissen hinsichtlich der Scanning Techniken von Hubschrauberpiloten spielt die Blickposition bei der Mensch-Maschine-Interaktion eine zunehmende Rolle (s. Exkurs unten). Hintergrund dessen ist die Tatsache, dass Operateure dann besonders effizient handeln, wenn sie Ziele oder Ereignisse, auf die sie reagieren sollen, in kürzester Zeit erfassen und ihre Reaktion direkt auf das System transferieren können. Kürzeste Übertragungsimpulse könnten in längerfristiger Zukunft über elektrische Hirnsignale des Operateurs realisiert werden (vgl. Wolpaw, Birbaumer, McFarland, Pfurtscheller, & Vaughan, 2002). Einen Schritt auf dem lang erwarteten Weg zur neuronalen Steuerung von technischen Systemen bietet die genauere Betrachtung der Augen- bzw. Kopfposition, da die Blickrichtung als zentraler Indikator für eine intendierte Handlung betrachtet wird. Inwiefern der Sinnspruch „*If Looks Could Kill*“ (Jönsson, 2005, S. 1) im militärischen Bereich bereits zur erfolgreichen Anwendung kommt, ist einer freien Recherche nur bedingt zugänglich.

Head Tracking zur Steuerung von Flugsystemen

Im Kampf(-Unterstützungs-)Hubschrauber Tiger ist es bereits möglich, über ein Pilotenvisioniksystem (montiert in der Hubschraubernase) und einem binokularen Helmsichtgerät (s. Abbildung 4) ein Wärmebild zu erzeugen sowie Luft-Luft-Raketen einzusetzen, welche über Head Tracking steuerbar sind. Die Erfassung der Koordinaten der Kopfposition im Verhältnis zur Lage des Hubschraubers und des Ziels im Raum erfolgt mithilfe eines elektromagnetischen Feldes im Cockpit. Die Ausrichtung des Hubschraubers zum gezielten Waffeneinsatz wird dabei parallel zur Kopfbewegung – und damit zur Blickrichtung – des Piloten realisiert.



Abbildung 4: Integriertes Helmsichtsystem⁴.

⁴ Links: http://farm6.static.flickr.com/5146/5639725353_f70017792d.jpg [Zugriff: 04.06.2013].
Rechts: http://farm6.staticflickr.com/5145/5639725341_e21ae5c5a8_z.jpg [Zugriff: 04.06.2013].

Der Fehlerradius bzw. die Effizienz beim Waffeneinsatz können hier nicht belegt werden, aber aus Sicht der Operateure wird von einer leichten Latenz⁵ zwischen Kopfbewegung und Ausführung durch das Sichtsystem berichtet. Im Gegensatz dazu ist bisher gut belegbar, dass beispielsweise PC-Nutzer, die blickrichtungsge- steuerte Videospiele spielen, von einem schnelleren, kontrollierteren, akkurate- ren und intuitiver bedienbaren Waffeneinsatz berichten als konventionelle Maus- oder Joystick-Spieler (Jönsson, 2005). An dieser Stelle ist davon auszugehen, dass Ergebnisse zum Lernen von Scanning Techniken in Videospiele auf den Kontext der Ausbildung von Hubschrauberführern übertragen werden können. So bildet im militärischen Bereich die Leistung in Videospiele einen zentralen Prädiktor und Trainingsbeitrag für den Einsatz im realen Szenario (Serious Gaming). Popescu, Romero und Usart (2012) liefern einen Überblick über die umfangreichen An- wendungsfelder von Videospiele allein im militärischen Ausbildungsbereich. In einer Metaanalyse von 38 Studien zeigen Wouters, van Nimwegen, van Oostendorp und van der Spek (2013), dass Videospiele im Vergleich zu konventio- nellen Methoden das Lernen ($d = 0.29$, $p < .01$) und das Behalten von Gelerntem ($d = 0.36$, $p < .01$) stärker fördern können.

Flugschüler, die für das Waffensystem Tiger vorgesehen sind, lernen den Waffen- einsatz mit Helmsichtsystem bisher nur über die verbale Kommunikation mit dem Fluglehrer. Problematisch ist hierbei, dass keine Videodatensätze mit Scanning Techniken auswertbar sind, sodass Fluglehrer zwar wissen, dass ein spezieller Fehler gemacht wurde, der genaue Kontext des Fehlers jedoch nicht nachvollzieh- bar ist. In Zukunft sollen Videosequenzen der Helmsicht gespeichert werden und in Debriefings (nach dem Flug) genauere Analysen stattfinden⁶. Allerdings können hiermit die tatsächlich getätigten Blicke (Nutzung der Symbolik im Helmsichtsyste- m, Fixierung eines Ziels) immer noch nicht konkret ermittelt werden. Eye Tra- cking Rückmeldungen könnten hier zukünftig eine sinnvolle Ergänzung darstellen, weil vom Flugschüler im Tiger spezielle, Scanning-bezogene Verhaltensweisen erlernt werden müssen. Da die Sichtprojektion auf die Okulare begrenzt ist, müs- sen beispielsweise Blickbewegungen außerhalb der Okulare vermieden, aber Kopfbewegungen vermehrt ausgeführt sowie in ihrer Geschwindigkeit adaptiert werden, um den Bildaufbau im Sichtfeld synchron zur Kopfbewegung zu ermögli- chen. Außerdem muss aufgrund der 2D-Projektion im Infrarot-Modus eine spezifi- sche Entfernungseinschätzung erlernt werden, die sich von der Orientierung an

⁵ Hauptmann Grebe, Modification Manager für UH-Tiger bei der Wehrtechnischen Dienststelle 61 der Bundeswehr; persönliche Kommunikation am 14.12.2011.

⁶ Oberstleutnant Fuchs, Testpilot UH-Tiger bei der Wehrtechnischen Dienststelle 61 der Bundes- wehr; persönliche Kommunikation am 22.11.2012.

realen, dreidimensionalen Reizen unterscheidet⁷. Damit Eye Tracking aber effektiv in die Ausbildung integrierbar wird, wäre speziell für diesen Anwendungsfall enorme technische Entwicklungsarbeit notwendig. Sinnvoll wäre hier zum Beispiel ein integrales Eye Tracking System am Helmvisier, da eine externe Kamerabrille, die dem Stand der Technik entspricht, das Helmsichtsystem überfrachten würde.

Ein Grund dafür, warum die Blicksteuerung von Mensch-Maschine-Systemen noch wenig verbreitet, aber seit Jahrzehnten beforscht ist, liegt in ihrer schwierigen technischen Umsetzbarkeit. Herausforderungen treten zum Beispiel dann auf, wenn Operateure ein Objekt unintendiert ansehen und es auf den ersten Blick aktivieren, weil keine Flächen in der Benutzerschnittstelle vorgesehen sind, auf denen der Blick ruhen kann (Midas Touch Problem; Nilsson, Gustafsson, & Carleberg, 2009). Darüber hinaus sind bisher komplexe Steuergesten (Drag & Drop, also Ziehen und kontrolliertes Loslassen eines Cursors) und eine vergleichbare Unterscheidung in rechte und linke Maustaste nur schwer realisierbar. Abhilfe wird bislang häufig in Form von virtuellen Tasten außerhalb von Displays bzw. Bildschirmen geschaffen (Schäfer, 2007) oder auf reines Head Tracking zurückgegriffen (s. oben). Ein anderes Problem der Generalisierbarkeit von Scanning Techniken wird dann relevant, wenn zum Beispiel Pilot-Assistenzsysteme Periodizitäten in Scanning Techniken verschiedener Piloten erkennen und für Entscheidungs- und Handlungsalgorithmen nutzen sollen (selbstadaptive Strategien von Operateuren; Donath, Rauschert, & Schulte, 2010). Doch am Anfang dieser relativ hohen Problemebene stehen basale Fragen der Blickregistrierung: 1. In welchen Fällen ist ein Blick oder eine Kopfposition Vorstufe einer Absicht? 2. Welcher Schwellenwert der Blicklänge ist hierfür aussagekräftig? 3. Wie kann man die Scanning Techniken eines Hubschrauberführers geeignet operationalisieren?

1.3 Impetus der Arbeit

Die Analyse von Scanning Techniken geschieht in bisherigen empirischen Arbeiten meist unter deskriptiver Angabe von Blickhäufigkeiten auf bestimmte Punkte sowie anhand von Einzelfällen und ist weder kreuzvalidiert (anhand z. B. eines subjektiven Fragebogens) noch in Form von Schwellenwerten für eine konkrete Blickdauer erforscht. Darüber hinaus liegen für die deutschen fliegenden Streitkräfte bisher noch gar keine Erkenntnisse hinsichtlich der Scanning Techniken von Piloten vor. Mit der vorliegenden Arbeit sollen diese Defizite bearbeitet werden. Die Studie möchte einen Vorschlag dafür liefern, wie unter Berücksichtigung wissen-

⁷ Oberstleutnant Fuchs und Technischer Oberregierungsrat van Schwartzberg, Projektgruppe UH-Tiger bei der Wehrtechnischen Dienststelle 61 der Bundeswehr; persönliche Kommunikation am 22.11.2012.

schaftlicher Kriterien eine quantitative und inferenzstatistische Analyse der Scanning Techniken von Hubschrauberpiloten durchgeführt werden kann. Damit soll auf der einen Seite geklärt werden,

- auf welche Weise sich die Scanning Techniken von erfahrenen Hubschrauberpiloten von denen unerfahrener Luftfahrzeugführer unterscheiden und
- welche Nützlichkeit Eye Tracking am Arbeitsplatz Simulatorcockpit aufweist, wenn Expertenmeinungen von Hubschrauberführern zugrunde gelegt werden.

Auf der anderen Seite soll damit ein weiterer Baustein für die Standardisierung der Eye Tracking Methode sowie für die Analyse der Scanning Techniken von Piloten gesetzt werden. Vor diesem Hintergrund wird das Verfahren an wissenschaftlichen Gütekriterien reflektiert (vgl. Rubio, Diaz, Martin, & Puente, 2004):

- Eye Tracking sollte zum einen *sensitiv* sein, das heißt es muss Unterschiede zwischen Anwendungssituationen bzw. Expertisegruppen aufdecken können. Weiterhin muss es *diagnostisch* sein, das heißt es muss Kausalfaktoren für Unterschiede erkennbar machen. Eye Tracking Daten sollen ferner *valide* sein, das heißt Messwertunterschiede müssen auf intendierte Versuchsbedingungen zurückzuführen sein und nicht auf konfundierte Störfaktoren. Die Messwerte sollten darüber hinaus *Reliabilitätsanforderungen* genügen, das heißt konsistent und wiederholbar abgebildet werden können. Die Methode sollte außerdem *nicht-intrusiv* sein, also typische Handlungsabläufe eines Operators nicht interferieren. Der Einsatz eines Eye Tracking Systems muss *ökonomischen* Ansprüchen genügen, indem sowohl Zeit- als auch Kostenaufwand kalkulierbar und verlässlich sind. Nicht zuletzt sollte Eye Tracking *Augenscheinvalidität* besitzen, so dass seine Nützlichkeit für die Operateure in der konkreten Anwendungssituation erkennbar ist.

Die Güte des Eye Tracking Verfahrens nach oben genannten Kriterien wird anhand einer praktischen Fragestellung aus der Heeresfliegerwaffenschule Bückeburg (Niedersachsen) untersucht. Im Vordergrund der Betrachtung stehen dabei Scanning Techniken, die von einsatzerfahrenen Hubschrauberpiloten an ihre Flugschüler vermittelt werden sollen. Einen Teil der Scanning Techniken im Hubschraubercockpit stellen dabei so genannte Target Fixations dar, also Fixierungen des Blickes auf bestimmte Objekte der Außenwelt oder auf Instrumente im Cockpit. Folgt man Experteninterviews, setzen Fluglehrer Target Fixations gezielt und intendiert ein und profitieren von ihrem peripheren Sichtfeld. Dagegen neigen unerfahrene Piloten häufiger zu unsystematischen und dysfunktionalen Scanning

Techniken, die ein umfassendes Situationsbewusstsein stören können und die ihnen teilweise aufwändig abtrainiert werden müssen (z. B. durch Abschalten der Cockpitinstrumente, um einen Blickwechsel zu erzwingen).

Scanning Techniken im Rahmen der Hubschrauberführergrundausbildung (HGA) erscheinen geeignet, die Sensitivität und diagnostische Güte der Eye Tracking Methode zu prüfen. Basierend auf der genannten Motivation entstand das hier beschriebene Dissertationsprojekt in Kooperation des Flugmedizinischen Instituts der Luftwaffe und der Heeresfliegerwaffenschule Bückeburg. Untersucht wird in dieser Arbeit konkret, inwiefern Scanning Techniken, subjektive Anforderungen bei einem Simulationsflug und erzielte Leistung im Training zwischen Fluglehrern und Flugschülern sowie zwischen unterschiedlichen Missionen differieren (Sensitivität vs. diagnostische Güte). Mithilfe eines Experimentalplans und detaillierter Abstimmung mit dem Ausbildungsplan im Simulatorzentrum soll die Validität und Reliabilität der Datenerhebung sichergestellt werden. Die Nützlichkeit der Methode für die Analyse von Scanning Techniken unter Berücksichtigung ihrer ökonomischen Erfordernisse wird durch eine Fragebogenerhebung mit den betroffenen Piloten bewertet.

In den folgenden Kapiteln werden zunächst technische, theoretische und empirische Hintergründe zu Eye Tracking, zur menschlichen Informationsverarbeitung und zum Ressourcenmanagement im Cockpit vorgestellt (Kapitel 2). Im Anschluss werden aus den Erkenntnissen und Überlegungen Hypothesen zu Scanning Techniken von Piloten abgeleitet und die vermuteten Zusammenhänge entsprechend operationalisiert sowie die zugehörigen Messmethoden und das empirische Vorgehen zur Hypothesenprüfung vorgestellt (Kapitel 3). Die Auswertung, Analyse und Darstellung der erhobenen Daten befindet sich in Kapitel 4. Im letzten Kapitel 5 werden die Erkenntnisse diskutiert, hinsichtlich ihrer wissenschaftlichen Güte bewertet und entsprechende Implikationen abgeleitet.

2 Technische, theoretische und empirische Hintergründe

2.1 Technische Grundlagen zu Eye Tracking

2.1.1 Funktionsweise von Eye Tracking Systemen

Die derzeit gängigen Eye Tracking Systeme messen den Blickpunkt durch **corneale⁸ Reflexion** (Poole & Ball, 2000). Wenn Infrarotlicht auf das Auge fällt, wird es auf der Netzhaut gespiegelt und tritt in Form von Reflexionen auf der Cornea wieder aus dem Auge aus. Die Punkte, die beim Austreten auf der Cornea erscheinen, werden **Purkinje-Reflexe** genannt. Der Reflexionsstrahl des Lichts wird aufgrund der Augenanatomie mehrfach gebrochen: Jeweils an Vorder- und Rückseite der Cornea (erster und zweiter Purkinje-Reflex) sowie an Vorder- und Rückseite der Linse (dritter und vierter Purkinje-Reflex, s. Abbildung 5).

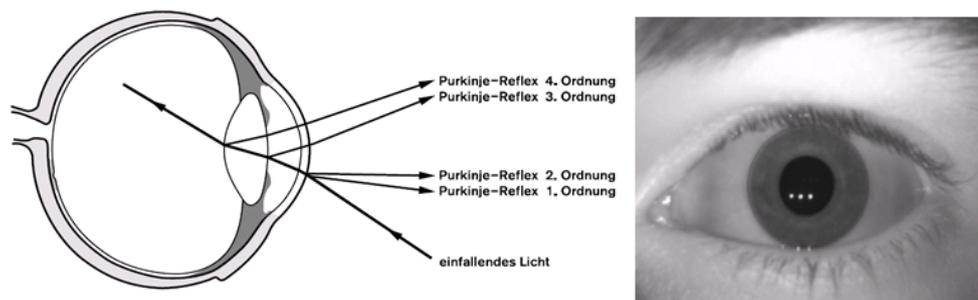


Abbildung 5: Prinzip der cornealen Reflexion und Purkinje-Reflexe
(übernommen aus Priese, Schmitt, & Lemke, 2007).

Die Verwendung von Infrarotlicht ist der Tatsache geschuldet, dass eine direkte Lichteinstrahlung auf das Auge die visuelle Wahrnehmung beeinträchtigen würde (Poole & Ball, 2000). Je höher die Ordnung des Purkinje-Reflexes, desto schwächer ist er sichtbar und desto schwerer ist er detektierbar. Der Reflex vierter Ordnung besitzt z. B. nur etwa ein Hundertstel der Helligkeit des ersten Reflexes. Hinzu kommt, dass sich die Reflexe teilweise verdecken können (Priese, Schmitt, & Lemke, 2007). Die Autoren (ebd., S. 7) erklären weiter: „Wenn bei einem fest auf einen Punkt fixierten Auge die Lichtquelle bewegt wird, dann bewegt sich der Purkinje-Reflex erster Ordnung gleichsinnig mit. [...] Da die Reflexe vierter Ordnung außerhalb von Laborbedingungen jedoch nicht zuverlässig gefunden werden können,

⁸ Die Cornea (dt. Hornhaut) ist der gewölbte vordere Teil der äußeren Augenhaut.

werden [...] [häufig] nur die Reflexe erster Ordnung verwendet [...].“ Auf die Blickrichtung wird aus dem Abstand der Reflexpunkte zur Pupille geschlossen (Sprenger, 2008). Poole und Ball (2000, S. 2) erklären weiter: „Once the image processing software has identified the centre of the pupil and the location of the corneal reflection, the vector between them is measured, and, with further trigonometric calculations, point-of-regard can be found“. Der so ermittelte Blickpunkt wird dann in ein zweidimensionales Koordinatensystem projiziert und beispielsweise als Fadenkreuz in einem Blickvideo (s. Abbildung 6) sichtbar.

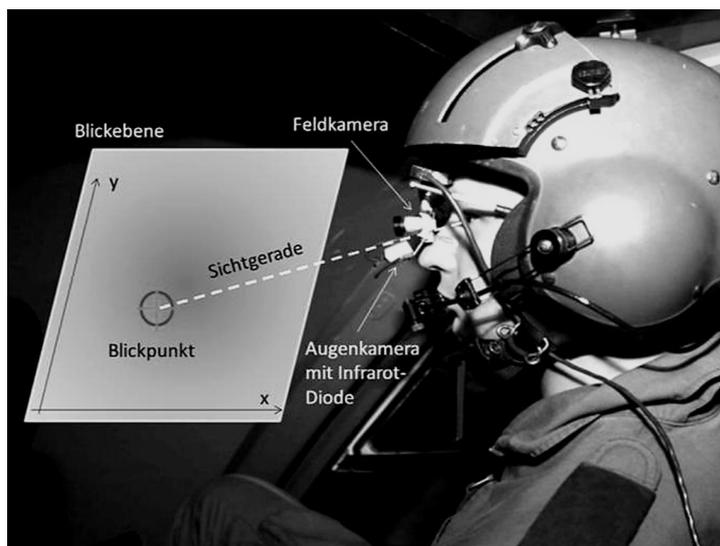


Abbildung 6: Projektion des Blickpunktes auf die Blickebene ($X, Y \in \mathbb{R}^2$; X_L, Y ; eigene Fotografie).

Typischerweise besteht ein Eye Tracking System aus einem Eye Tracker, das heißt aus einer Kamera, welche die Blickdaten aufzeichnet, einer Feldkamera, die den Untersuchungsraum filmt und einem PC, der die Daten speichert. Für eine optimale Datenanalyse sollte außerdem eine Software vorhanden sein, die es ermöglicht Interessenbereiche (**Areas of Interest; AOI**) im Untersuchungsraum festzulegen und darauf bezogene Scanning Techniken auszuwerten. Unter einer AOI wird der abgegrenzte Bereich im Untersuchungsraum verstanden, in dem Scanning Techniken für die Fragestellung analysiert werden sollen. Steht z. B. die visuelle Nutzungshäufigkeit eines Head Up Displays (HUD) im Fokus, dann bildet das HUD ein AOI. Idealerweise ist es mit der Software auch möglich, Scanning Techniken grafisch darzustellen bzw. Blickvideos zu schneiden. Ein zentrales Element jeder Eye Tracking Software ist ein Modul zur Kalibrierung der Augenkamera. Die gängigen Softwares verfügen über eine automatische Pupillenerkennung, die vom Experimentator auch manuell reguliert werden kann. Eine gründliche Kalibrierung gewährleistet valide Blickdaten, kann aber sehr zeitaufwändig sein (Sprenger, 2008, berichtet von bis zu zwei Stunden pro Proband).

ler Abhängigkeit des Abstandes zwischen Blickebene und Kalibrierradius. Sie findet ihre Quantifizierung in der klassischen Winkelformel der Trigonometrie:

$$\tan(\alpha) = \frac{\text{Gegenkathete}}{\text{Ankathete}}$$

Wird die Kalibrierung im Flugsimulator auf der Ebene der Frontscheibe des Hubschraubers vorgenommen, beträgt der Kalibrierradius etwa $r_1 = 100$ cm (Ankathete). Bei $\alpha = 0.5^\circ$ Messgenauigkeit beträgt die Maximalabweichung des gemessenen vom tatsächlichen Blickpunkt nach Umstellung der Formel 0.87 cm, welche dem Stand der Technik der marktverfügbaren Eye Tracking Systeme entspricht. Zur Messgenauigkeit von 0.5° bemerkt Poitschke (2011, S. 37): „Verglichen mit der menschlichen Refixationsgenauigkeit von ca. 1° [...], stellt dies einen ausreichend guten Wert dar.“ Ausgehend davon, dass die Projektionsfläche des EC 135 Simulators etwa 300 cm vom Auge des Piloten entfernt ist (s. Abbildung 8), erhöht sich die Rekonstruktionsungenauigkeit des Blickpunktes proportional auf 2.62 cm.

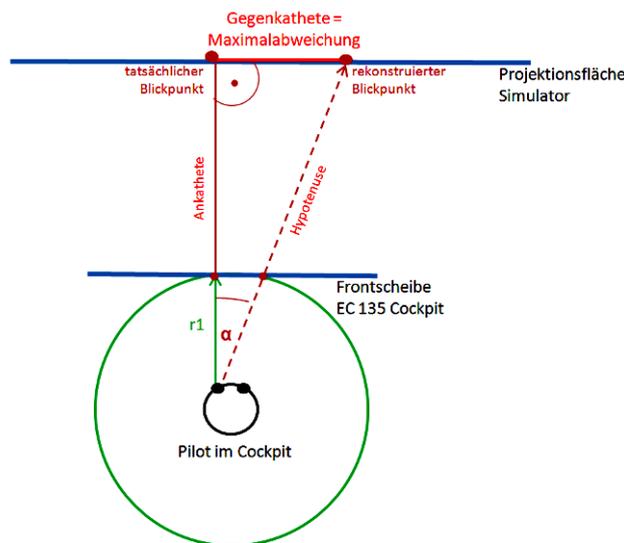


Abbildung 8: Trigonometrisches Berechnungsprinzip der Messgenauigkeit im Flugsimulator¹¹.

Aktuell existieren Entwicklungsversuche für die **Kalibrierung auf weiteren Tiefenebenen**, beispielsweise für Head Mounted Eye Tracking in virtuellen oder realen Umgebungen¹². Um präzise Eye Tracking Ergebnisse in 3D-Umwelten zu erhalten, werden teilweise binokulare Messverfahren eingesetzt (z. B. Duchowski, Medlin, Gramopadhye, Melloy, & Nair, 2001). Der Kerngedanke hierbei liegt in der Zuhilfenahme des interpupillaren Abstandes und damit der Konvergenzstellung der

¹¹ Eigene Skizze; nicht maßstabsgetreu.

¹² Johann Moor, Ergoneers GmbH, persönliche Kommunikation am 31.01.2013.

Augen, da diese die Querdissparation¹³ und das räumliche Sehen moduliert. Aus dem Konvergenzwinkel der Augen kann so auf die Tiefenebene des Blickpunktes geschlossen werden (Prinzip s. Abbildung 9, links). Duchowski et al. (ebd.) legen einen Machbarkeitsversuch mit ersten Blickdaten in einer realen Umgebung vor, kommen aber zu dem Schluss, dass die Ergebnisse noch mit Rauschen belastet sind und die tatsächliche Anzahl an Fixationen unterschätzt wird. In einem anderen Experiment von Pfeiffer, Donner, Latoschik und Wachsmuth (2007) wurde die Genauigkeit der Messung von Blickpunkten in einem virtuellen Raum (PC-assistiert) mit bis zu 30 cm Tiefe überprüft. Verglichen wurde die Messgenauigkeit der Blickpunkte mit Remote Eye Tracking bei der klassischen Kalibrierung auf einer 2D-Ebene vs. einer Kalibrierung mit einem 3D-Würfel (s. Abbildung 9, rechts).

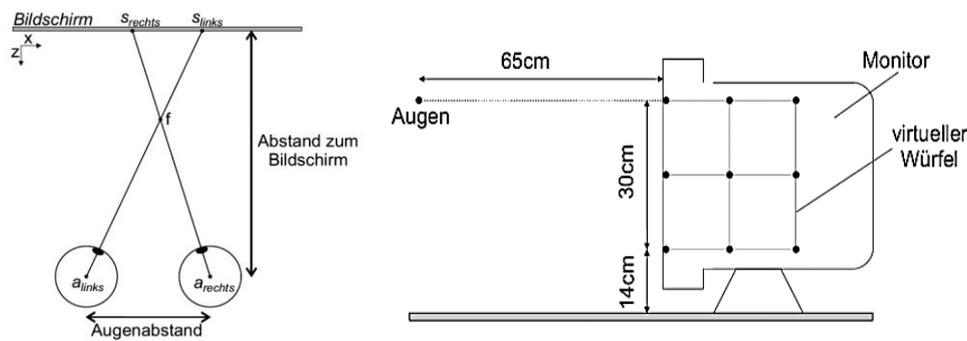


Abbildung 9: Versuchsaufbau bei Pfeiffer et al. (2007; übernommen aus ebd.).

Im Resultat zeigte sich, dass die 3D-Kalibrierung im Vergleich zur klassischen sogar zu schlechteren Ergebnissen führte (42 % vs. 75 % korrekt identifizierter Blickpunkte). Nur in kritischen Bereichen, in denen virtuelle Objekte direkt hintereinander lagen, verbesserte sich die Erkennungsrate in der 3D-Bedingung im Vergleich zur klassischen Kalibrierung (42 % vs. 30 % korrekt identifizierter Blickpunkte). Solche visuellen Ambiguitäten betreffen in der Luftfahrt z. B. Blicke auf ein HUD vs. Blicke in den Luftraum direkt hinter dem HUD. In der vorliegenden Untersuchung hingegen sind die Ebenen OTW (Out The Window) und Blicke ins Cockpit auf 2D-Ebene trennscharf und ohne Überlappungen zu unterscheiden. Entsprechend ist eine 3D-Kalibrierung nicht notwendig, wenngleich unter Berücksichtigung des Stands der Technik bei weitwinkligen Blicken in die Außenwelt die oben quantifizierten Messgenauigkeiten in Kauf genommen werden müssen.

¹³ Leicht versetzte Abbildungen auf der Netzhaut, die räumliches Sehen ermöglichen, nennt man Querdissparation.

2.1.3 Arten von Eye Tracking Systemen

Eye Tracking Systeme werden in zwei Arten unterteilt: **Head Mounted Systeme** und **Remote Eye Tracker**. Erstere werden von Probanden in Form von Brillen, Helmen oder Stirnbändern auf dem Kopf getragen, wobei sich die Augenkamera wenige Zentimeter vor einem Auge befindet (s. Abbildung 10).

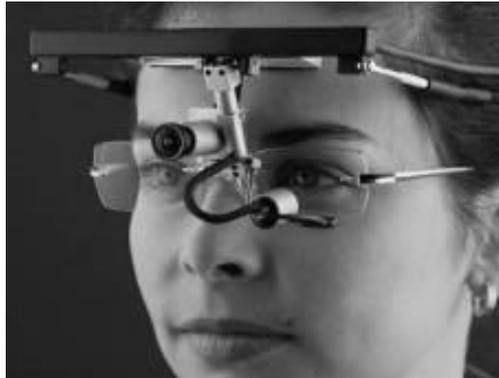


Abbildung 10: Head Mounted System¹⁴.

Remote Eye Tracker hingegen sind z. B. am PC fest stationiert und kommen nicht direkt mit der Testperson in Berührung (s. Abbildung 11).



Abbildung 11: Remote System (übernommen aus DeSantis et al., 2005).

Da Head Mounted Kamerasysteme direkt am Kopf der Probanden befestigt werden, können Scanning Techniken dynamisch, das heißt auch bei z. B. Umherlaufen oder starken Kopfdrehungen, aufgezeichnet werden (vgl. DeSantis, Zhou, & Ramey, 2005). Der Vorteil von Remote Systemen liegt dagegen in ihrer fixen Stationierung unabhängig vom Probanden, wodurch keine Unannehmlichkeiten oder Bedarf an Nachkalibrierung entstehen. Ein Remote Eye Tracker, der komplett stationär ist und nicht über einen Head Tracker verfügt, erfordert allerdings eine

¹⁴ <http://www.ergoneers.com/de/products/dlab-dikablis/overview.html> [Zugriff: 28.04.2013].

Fixierung des Kopfes (z. B. über Kinnstütze; DeSantis et al., 2005). Nachteilig bei Head Mounted Eye Trackern ist die Tatsache, dass sie Kopfschmerzen und Druckstellen auslösen können und insbesondere bei längeren Versuchen zu Unannehmlichkeiten führen. Damit können sie als störend wahrgenommen und zu Artefakten im Versuch werden.

Dennoch gelten Head Mounted Eye Tracking Systeme als **messgenauer** im Vergleich zu Remote Systemen (Sprenger, 2008), denn speziell beim Remote Eye Tracking entstehen diffuse Reflexionen (Rauschen) nicht nur aufgrund von Augenbewegungen, sondern auch bei Kopfbewegungen von Probanden. Damit können die Purkinje-Reflexe artefaktbelastet sein. Diffuse und nicht präzise erfassbare Reflexionen können z. B. bei bestimmten Kopf-Auge-Positionen generiert werden, wo Wimpern die Reflexion stören oder die Pupille fälschlicherweise zwischen Cornea und Sclera erkannt wird (Hennessey & Lawrence, 2008). Poitschke (2011, S. 75) fasst die Ursachen von größeren Messungenauigkeiten beim Remote Eye Tracking zusammen und betont die Rolle des Abstandes zwischen Auge und Eye Tracker: *„Aus der größeren Entfernung zur Kamera und den Beleuchtungseinheiten rühren eine geringere Größe und ein geringerer Abstand der Reflexionen (Glints) im Videobild, weshalb sich deren exakte Position schwerer bestimmen lässt. Weiterhin können bei zu großer Distanz zum Tracker die beiden Glints im Videobild mehr und mehr zu einem Blob ‚verschmelzen‘. Diese Faktoren führen zu einem schlechteren Konvergenzverhalten der Rekonstruktionsalgorithmen. Zusätzlich wirkt sich die wachsende Distanz drastisch auf die Bildqualität aus. Als letzter Punkt sei die geringere örtliche Auflösung des Auges im Videobild angeführt, welche sich in größeren Distanzen negativ auf die Ergebnisse der Parameterextraktion auswirkt.“*

Hennessey und Lawrence (2008) führten ein Experiment durch, bei dem sie 189 verschiedene Kopf- und Augenpositionen von sieben Personen hinsichtlich der Messgenauigkeit der Blickpunkte mit Remote Eye Tracking verglichen. Die Autoren fanden heraus, dass die Schätzung des Blickpunktes dann am genauesten gelang, wenn der Zentroid aller Purkinje-Bilder als Referenzwert der cornealen Reflexion gebildet wurde. Wenn die Blickpunktbestimmung lediglich aufgrund eines Purkinje-Bildes vorgenommen wurde, scheiterte sie in bis zu 27 % der Fälle. Wenn also beim Remote Eye Tracking eine akzeptable Pupillenreproduktionsgenauigkeit trotz Kopfbewegungen erzielt werden soll, sollten die Corneareflexe aggregiert werden, was mit Informationsverlusten einhergehen kann. Direkte Äquivalenz-Untersuchungen zu Remote vs. Head Mounted Eye Tracking liegen zum Kenntnisstand der Autorin nicht vor. Welche Art von Eye Tracking System für einen Versuch angewendet wird, hängt daher von der individuellen Fragestellung ab. Eye Tracking Analysen am PC beispielsweise – also einer wenig variablen bzw. wenig

dynamischen Arbeitsumgebung, in der Probanden sehr nah am Interessenbereich des Experimentators sitzen – eignen sich für die Verwendung von Remote Systemen. Untersuchungen, die in einer dynamischen Umgebung stattfinden sollen, wie z. B. im Flug- bzw. Fahrsimulator oder im Kaufhaus zur Bestimmung von Konsumentenverhalten, erfordern Head Mounted Systeme.

2.1.4 Heat Maps

Beiden Arten von Eye Tracking Systemen ist gemein, dass die Softwares ein prominentes grafisches Auswertungstool, die **Heat Map**, bieten. In Abbildung 12 ist eine Heat Map eines Kaufhausregals dargestellt, die solche AOIs mit einer dunkleren Farbe kennzeichnet, auf die durch Konsumenten häufiger geblickt wurde (Datenberechnung mit der Eye Tracking Software). Anhand dieser Messung könnte beispielsweise der Rückschluss gezogen werden, dass Produkte, die umsatzstärker sind, in der Mitte des Regals platziert werden sollten, da sich hierin die Blicke der Personen häuften (rot dargestellte Flächen).



Abbildung 12: Heat Map¹⁵.

Allerdings ist die Interpretation einer solchen Heat Map mit Vorsicht zu tätigen. Schäfer (2007, S. 26 und S. 28) erklärt hierzu: „[Für die Heat Map] wird ein Integer Array¹⁶ mit der Größe der Bildschirmauflösung angelegt und jedem Pixel eine Zelle zugeordnet. Nach der Initialisierung ist die komplette Matrix mit Nullen gefüllt. Streift der Blick einen bestimmten Pixel, wird der Wert in der zugehörigen Zelle um eins erhöht. Dabei kommt wieder das Problem der räumlichen Ausdehnung einer Fixation ins Spiel. Würde nur genau die zur Blickkoordinate passende Zelle iteriert, sähe die Karte am Schluss meistens sehr leer aus. Die Wahrscheinlichkeit, dass ein Pixel wiederholt getroffen wird, wäre zu gering um aussagekräftige Bilder mit mehreren Farben zu erhalten.“ Der Autor führt weiter aus, dass

¹⁵ http://www.rogil.be/images/heatmap_c.gif [Zugriff: 28.04.2013].

¹⁶ Englisch für eine ganzzahlige Datenmatrix.

deswegen nicht nur der betreffenden Koordinate bei einer Fixation der Wert 1 zugewiesen werde, sondern auch weiteren, umliegenden Punkten, denen aber aufgrund ihrer Vielzahl nicht vollkommen differente Farben zugeordnet werden können. Entsprechend sind der Umkreis des eigentlichen Fixationsortes und die Farbgebung der Heat Map eher unscharf. Das grafische Werkzeug ist daher aus wissenschaftlicher Perspektive zu unpräzise skaliert für die exakte Analyse von Scanning Techniken. Will man einem Einzelfall visuelle Rückmeldung über Scanning Techniken ermöglichen (s. Abschnitt 4.1.3), mag eine Heat Map zwar nützlich sein, zur quantitativen Auswertung ist sie gänzlich ungeeignet.

2.1.5 Überblick der Nach- und Vorteile von Eye Tracking

Unter Berücksichtigung der erwähnten Unterschiede zwischen Remote und Head Mounted Systemen sowie den Besonderheiten von Heat Maps wird das Eye Tracking Verfahren als solches hinsichtlich seiner methodischen Güte kontrovers diskutiert. Die **Nachteile** des Verfahrens lassen sich im Überblick wie folgt zusammenfassen (vgl. z. B. Schandry, 1989, S. 111f; Poole & Ball, 2000; Colvin, Dodhia, Bechler, & Dismukes, 2003; Chin, Lee, & Ramey, 2005; Monk & Kidd, 2007; Sprenger, 2008; Metz, 2009; Geise 2012):

- Die Kalibrierung kann sehr zeitaufwändig sein.
- Derzeit ist die Kalibrierung nur für zweidimensionale Ebenen technisch ausgereift; damit müssen Messungenauigkeiten außerhalb des Kalibrier-radius in Kauf genommen werden.
- Es kann zur Interferenz der Apparatur mit der eigentlichen Aufgabe des Probanden und damit zur Entstehung von Artefakten kommen.
- Es können Messprobleme bei Probanden mit Kontaktlinsen, Brillen, starker Schminke oder sehr langen Wimpern auftreten.
- Es können Aufmerksamkeitsartefakte durch starke Kontraste bzw. grafische Pop Outs sowie eine hohe Objektdichte entstehen.
- Es existiert bisher keine klar definierte zeitliche Untergrenze für Fixationen¹⁷.
- Es können unabhängig vom Experimentalparadigma keine eindeutigen Rückschlüsse von Blicken auf tatsächlich wahrgenommene Informationen gezogen werden¹⁸.
- Die Interpretation von Heat Maps ist eher unscharf.

¹⁷ S. vertiefend Abschnitt 2.2.6.

¹⁸ S. vertiefend Abschnitt 2.2.6.

- Es existiert die Möglichkeit der Datenmanipulation durch die Experimentatoren (Post-Hoc-Eingriffe in Blickvideos).
- Die Ergebnisse lassen sich selbst innerhalb gleicher Forschungsbereiche nur schwer generalisieren (z. B. Scanning Techniken im Kampfjet vs. Hub-schrauber), da kaum standardisierte experimentelle Ansätze existieren.

Geise (2012, S. 25) reflektiert zusammenfassend: *„Das besondere Problem der Interpretation von Blickverlaufdaten liegt, wie bei fast allen apparativen, rezeptionsbegleitenden Beobachtungsverfahren [...], darin, dass der Forscher in den meisten Rezeptionssituationen passiver Beobachter ist und damit nur teilweise nachvollziehen kann, was der Rezipient gerade tut, warum er dies tut und was er dabei denkt. Zudem ist gerade die Interpretation psychophysiologischer Daten oft widersprüchlich.“*

Andere Autoren vertreten Eye Tracking als *„widely accepted and valid measure of the visual demand [...]“* (Ablassmeier, Poitschke, Wallhoff, Bengler, & Rigoll, 2007, S. 2251) und betonen die **Vorteile** (vgl. auch Chin et al., 2005; Tsai et al., 2007):

- Mit Eye Tracking kann eine Vielzahl physiologischer Größen bzw. Blickparameter erfasst werden¹⁹.
- Die Probanden müssen keine Selbstauskünfte geben, wodurch Eye Tracking in die Sparte der objektiven Messverfahren einzuordnen ist.
- Die Daten können online und ohne Unterbrechung einer Tätigkeit des Probanden aufgezeichnet werden.
- Neuere Eye Tracking Systeme erlauben die Datenaufzeichnung bei Probanden mit (Gleitsicht-)Brillen und Kontaktlinsen.
- Head Mounted Systeme erlauben die Datenaufzeichnung auch in dynamischen Umgebungen.
- Es lassen sich quantitative Daten extrahieren.
- Die Daten können mit Hilfe von Software und „per Mausklick“ ausgewertet werden.
- Eye Tracking Ergebnisse können leicht mit subjektiven Daten kombiniert werden.
- Die EN ISO 15007 Nr. 1 und 2 versuchen, Eye Tracking Studien zu standardisieren. Sie liefern eher knappe und auf den Automobilkontext bezogene Anhaltspunkte für den Versuchsaufbau von Eye Tracking Experimenten²⁰.

¹⁹ S. vertiefend Abschnitt 2.1.6.

²⁰ Inzwischen steht eine Neunormierung (EN ISO 15007-1 und ISO 15007-2) vor der Verabschiedung (Dr. Christian Lange, Ergoneers GmbH, am 22.03.2013 per Email-Kommunikation).

- Eye Tracking Befunde sind für viele verschiedene Forschungsbereiche von Bedeutung (z. B. Medienwissenschaften, Kognitionswissenschaften, Ingenieurwesen etc.)²¹.

Durch die technischen und methodischen Herausforderungen, die den Vorteilen der Methode gegenüber stehen, gestaltet sich das **Kosten-Nutzen-Verhältnis** einer Eye Tracking Studie in gewisser Weise **kritisch**. Dies erfordert eine detaillierte Abwägung auf Seiten der Experimentatoren. Wie den theoretischen und methodischen Defiziten im Rahmen dieser Studie begegnet wurde, wird ab Kapitel 2.2 und 3 näher erläutert. Zunächst werden die mit Eye Tracking messbaren Blickparameter vorgestellt.

2.1.6 Blickparameter und ihre Interpretation

Mit Eye Tracking können verschiedene Parameter erfasst werden, wobei die Anzahl der Blicke, die Blickdauer, die Anzahl der Sakkaden, der Blickweg und die räumliche Verteilung von Blicken zunächst einige Beispiele darstellen. Merchant (2001) unterteilt die Reaktionsmöglichkeiten der Augen ganz allgemein in Reaktionen mit und ohne Bewegung. Zu erstgenannten Reaktionen gehören Lidschlag, Akkomodation der Linse und die Pupillenreaktion. Zu den Letztgenannten zählt der Autor Sakkaden, Augenfolgebewegungen, Nystagmus, den Vestibulo-Okular-Reflex und Vergenzbewegungen. Schandry (1989, S. 275f) differenziert Letztere außerdem in Konvergenz- und Divergenzbewegungen. Das gesamte Blickfeld des Auges beträgt laut Schandry (1989) etwa 100°; Metz (2009) ergänzt, dass das reine foveale Sehen auf einen Ausschnitt von lediglich 2° begrenzt ist. Merchant (2001) erklärt, dass foveales Sehen etwa 3 bis 4 Buchstaben, parafoveales Sehen etwa 10 bis 15 Buchstaben bzw. 5° um die Fovea und peripheres Sehen alles Weitere im visuellen Feld umfasst²². Wie lassen sich aber die einzelnen Blickparameter differenzieren (vgl. Übersicht bei Poole & Ball, 2000)?

Fixationen dienen dazu, mithilfe fovealen, parafovealen und peripheren Sehens Informationen zu enkodieren. Eye Tracking misst dabei aber nur die fovealen Sehanteile direkt (corneale Reflexion); alle anderen (parafoveale und periphere) müssen indirekt erschlossen werden. Längere Fixationen können sowohl Indikator für hohes Interesse an einer Information sein als auch Indikator für hohe Komplexität und Enkodierungsschwierigkeiten durch die Versuchsperson (Poole & Ball,

²¹ S. vertiefend Abschnitt 2.1.8.

²² Strasburger, Rentschler und Jüttner (2011) geben an: Durchmesser der Fovea: 5.2°, Durchmesser des parafovealen Bereichs: 5 bis 9°, Durchmesser des perifovealen Bereichs: 9 bis 17°. Alle Bereiche zusammen ergäben die Macula (17°). Das zentrale visuelle Feld umfasse 60°, das periphere visuelle Feld umfasse 60 bis 180° horizontal.

2000; Geise, 2012). Diesbezüglich wird davon ausgegangen, dass Fixationen mit Aufmerksamkeitsprozessen assoziierbar sind: Kurz bevor Probanden einen bestimmten Punkt fixieren, lenken sie die Aufmerksamkeit auf diesen Punkt. Dieser heuristische Zusammenhang zwischen Aufmerksamkeit und Fixationen wird traditionell auch als Eye-Mind-Hypothese bezeichnet (Just & Carpenter, 1980; vgl. z. B. Poole & Ball, 2000; Tsai, Viirre, Strychacz, Chase, & Jung, 2007; Metz, 2009). Es wird weiterhin spekuliert, dass die **Fixationsdauer** als Indikator für die Verarbeitungsdauer einer visuellen Information betrachtet werden kann und „*dass die Verschiebung der Aufmerksamkeit rund 100 ms vor einer Sakkade stattfindet*“ (Metz, 2009, S. 14). Der Ort der Fixation ist jedoch nicht immer gleichzeitig der Ort der Aufmerksamkeit (s. Abschnitt 2.2.6). Bei beispielsweise Beobachtungen aus dem Augenwinkel (peripheres Sehen; z. B. Metz, 2009, S. 14), Starren auf einen Punkt bei gleichzeitiger Konzentration auf auditive Reize oder „Tagträumen“ lässt sich die Fixationsdauer nicht als valider Indikator für Aufmerksamkeit heranziehen. Metz (2009) weist darauf hin, dass für die meisten solcher Verarbeitungsprozesse eine willentliche bzw. bewusste Unterdrückung der mit der Aufmerksamkeitsverschiebung gekoppelten Blickbewegungen erforderlich ist.

Sakkaden treten dann auf, wenn der Blick von einem Objekt zum anderen wechselt (vgl. ISO 15007-1, 2002). Es handelt sich um ruckartige, sprunghafte, sogenannte ballistische Blickbewegungen (Schandry, 1989, S. 275), deren Richtung vor der Ausführung festgelegt ist und deren Geschwindigkeit etwa 700° pro Sekunde beträgt (Merchant, 2001). Während einer Sakkade wird keine visuelle Information verarbeitet. Allerdings können häufige, regressive Sakkaden ein Indikator für hohe Informationskomplexität sein (Poole & Ball, 2000). Während einer Fixation kommt es außerdem häufig zu Mikrosakkaden, die den physiologischen Nystagmus (s. unten) des Auges korrigieren (vgl. Merchant, 2001). Sakkaden werden von den meisten Eye Tracking Systemen nicht direkt erfasst, sondern es wird häufig über gemessene Blickwechsel hierauf geschlossen.

Mehrere aufeinander folgende Fixationen auf einen bestimmten Punkt oder ein AOI zur Informationsaufnahme lassen sich als **Blick** (auch: Gaze oder Glance) zusammenfassen (Hendee & Wells, 1993). In der ISO 15007-1 (2002, S. 5) wird die **Blickdauer** (auch: Dwell Time oder Gaze Duration) als „*Zeitspanne, beginnend mit dem Zeitpunkt, von dem an sich der Blick auf ein Ziel [...] hin bewegt, bis zu dem Moment, zu dem er sich wieder davon abwendet*“ definiert. Der Unterschied zur Fixationsdauer liegt darin, dass in der Blickdauer nicht nur die Verweildauer, sondern auch die Übergangszeit (also Zeit, bis der Blick auf dem Ziel gestoppt wird) enthalten ist. Geise (2012, S. 33) weist darauf hin, dass „*bei längerer Blickverweildauer ein höherer Grad der Informationsaufnahme wahrscheinlich [ist]; das gilt*

vor allem, da mit steigender Fixationsdauer von einer Zunahme der kognitiven Verarbeitungstiefe ausgegangen werden kann“. Ein theoretisches Defizit herrscht aber dahingehend, ob bei einer sehr langen Blickdauer, wie oben bei Fixationsdauer schon erwähnt, immer von tieferer kognitiver Verarbeitung die Rede sein kann, oder ob weitere Mechanismen zugrunde liegen.

Als **Blickweg** wird eine komplette Sakkade-Fixation-Sakkade-Sequenz bezeichnet (Poole & Ball, 2000). Blickwege als Ganzes werden häufig nur qualitativ und einzelfallartig analysiert und beschrieben, um einen Eindruck über die Scanning Techniken einer Person zu erhalten.

Aufgabentypische Sequenzen von Augenbewegungen werden als **Scanning Technik** oder **Scan Patterns** (Henderson, 2003) bezeichnet. Scanning Techniken schließen aufgrund von aufgabenrelevantem Wissen eine bestimmte Abfolge von Fixationen, Blicken und damit Blickwegen („*gaze-control policy or strategy relevant to a given task*“; ebd., S. 501) ein. Sie gelten für die visuelle Wahrnehmung von Landschaftsszenen aber als kaum erforscht. Insbesondere für Scanning Techniken in Glass Cockpits mangelt es an Erkenntnissen: *“However, no standardized method of scanning glass cockpit displays has been developed. With the integration of an eye tracker, future research may investigate how pilots scan [...]”* (Lindo, Deaton, Cain, & Lang, 2012, S. 70).

Augenfolgebewegungen werden dann ausgeführt, wenn ein bewegtes Zielobjekt verfolgt wird (Schandry, 1989, S. 275), z. B. bei Beobachtung eines Pendels. Augenfolgebewegungen gehen ab einer Objektgeschwindigkeit von 100° pro Sekunde in Sakkaden über, da das Auge dieser Geschwindigkeit nicht mehr konstant folgen kann. Sakkaden dienen während Augenfolgebewegungen außerdem als Korrekturen zwischen Auge und Zielobjekt (Merchant, 2001).

Als **Vergenzbewegungen** werden gegenläufige Blickbewegungen bezeichnet. Konvergenz tritt auf, wenn nahe Objekte fixiert werden, Divergenz hingegen erscheint bei Fixationen von weiter entfernten Objekten (Schandry, 1989, S. 275).

Der **Vestibulo-Okular-Reflex (VOR)** ist eine Anpassungsbewegung der Augen für den Fall, dass sich die Position des Kopfes im Raum ändert. Er wird über Signale der Bogengänge des Innenohrs ausgelöst (Schandry, 1989, S. 276). Bei Auftreten des VOR rotieren die Augen entgegengesetzt zur Kopfbewegung um Objektfixationen trotz Bewegungen des Beobachters zu ermöglichen (Merchant, 2001).

Augenzittern oder ruckartige Blickbewegungen, die beim Fixieren von Objekten auftreten, werden unter dem Begriff **Nystagmus** zusammengefasst. Physiologischer und in diesem Sinne endogener Nystagmus tritt beim Fixieren von Objekten

immer auf, hat eine Amplitude von maximal 1° und führt dazu, dass das fixierte Bild auf ständig wechselnde Punkte der Retina trifft. Pathologische Formen von Nystagmus können auf Erkrankungen des Hirnstamms bzw. Kleinhirns hinweisen (Schandry, 1989).

Unter **Lidschlag** wird die teilweise oder komplette Schließung des Auges durch das Lid verstanden (Merchant, 2001). Die **Lidschlussdauer** variiert zwischen etwa 40 bis 200 ms in Abhängigkeit der Aufgabe bzw. Stimuli (Horend, 2004). Die Anzahl der Lidschlüsse über einen bestimmten Zeitraum wird als **Lidschattenfrequenz** bezeichnet. Das Auge gilt als geschlossen, wenn weniger als 35 % der Pupille sichtbar sind (vgl. Horend, 2004).

Die Form der Linse des Auges wird über die Ziliarmuskeln gesteuert und determiniert die Fokussierung von Objekten sowie deren Repräsentation auf der Retina. Diese Anpassung der Linse wird als **Akkomodation** bezeichnet (Merchant, 2001; ISO 15007-1, 2002).

Die Weite der Pupille variiert mit der Helligkeit der Umgebung und regelt den Lichteinfall auf das Auge. Die **Pupillendilatation** (auch: Pupillenruhe) dient der Korrektur der Beleuchtungsstärke der Retina und wird über einen physiologischen Regelkreis gesteuert. Diese Fähigkeit des Auges wird auch mit **Adaption** umschrieben (z. B. ISO 15007-1, 2002). Die Weite der Pupille bzw. ihre Schwankungen als vielfach diskutierte Indikatoren für kognitive Prozesse, wie z. B. Satzverarbeitung, Gedächtnisleistungen und Aufmerksamkeit, werden in Abschnitt 2.1.7 erläutert, da die Pupillenreaktion im Regelfall nicht über klassische Eye Tracking Systeme gemessen wird. Vielmehr existiert hierfür ein eigenständiges Vorgehen; die Pupillometrie.

2.1.7 Abgrenzung zu anderen okulografischen Messmethoden

Neben Eye Tracking existiert eine Reihe weiterer physiologischer Verfahren, mithilfe derer auf kognitive Prozesse geschlossen werden soll. Die Instrumente werden im Folgenden kurz vorgestellt und mit Eye Tracking verglichen.

Über **Pupillometrie** wird die Pupillenweite bzw. Pupillendilatation gemessen, welche als „objektive physiologische Manifestation psychischer Phänomene“ betrachtet wird (Schandry, 1989, S. 279; Beatty, 1982). Einsatz findet diese Methode unter anderem bei der Differenzierung psychischer Störungen (z. B. Depression, substanzinduzierte psychische Störungen; s. Keller, 2004) oder bei der Detektion von Ermüdung (Poole & Ball, 2000; Keller, 2004; Regen, 2009) sowie Täuschung oder Lügen (z. B. Dionisio, Granholm, Hillix, & Perrine, 2001). Der Pupillenruheindex

(PUI), welcher als Maß für die Schwankung der Pupillengröße bei konstanten Lichtverhältnissen gilt, errechnet sich aus „*der Summe der Beträge der Differenzen des mittleren Pupillendurchmessers über die Zeit*“ (Regen, 2009, S. 16). Ältere empirische Arbeiten berichten Schwankungen der Pupille zwischen 0.1 – 0.7 mm in Abhängigkeit der Aufgabe (z. B. Satzverarbeitungs- oder Vigilanzaufgaben, für einen Überblick s. Beatty, 1982), wobei die Befunde inkonsistent sind. Weitere Parameter der Pupillometrie sind z. B. der durchschnittliche Durchmesser oder die Varianz des Durchmessers. Schandry, (1989, S. 280) bezeichnet die Varianz als am besten geeignet um „*mentale Konzentration*“ zu beschreiben (s. auch Klingner, Kumar, & Hanrahan, 2008). Vorteile des Indikators Pupillenweite sind das schnelle Ansprechen auf Veränderungen, die kontinuierliche Messbarkeit ohne dass Probanden Auskünfte geben oder Aufgaben erfüllen müssen und die Unverfälschbarkeit durch Versuchsteilnehmer (vgl. Schultheis, 2004). Die Methode weist allerdings einige messmethodische Nachteile auf. Problematisch an Daten der Pupillometrie ist, dass sich die Pupillenweite infolge von Lichtveränderungen ständig ändert. Messungen, die unter normalen Tageslichtbedingungen stattfinden, sind also per se mit Artefakten belastet und nicht fehlerfrei interpretierbar. Bei plötzlichem starkem Lichteinfall kann die Pupillenweite um bis zu 5 mm schwanken (Lichtreflex; Schultheis, 2004). Außerdem unterliegt die Pupillenweite spontanen, endogenen Schwankungen (Hippus; z. B. Schultheis, 2004), die die Messwerte mit Rauschen überdecken können. Einige Autoren raten deswegen zur Mittelung über viele Einzelmessungen, wobei Informationsverluste in Kauf genommen werden müssen (Schandry, 1989, S. 279ff.). Eine empirische Arbeit von Klingner et al. (2008) analysiert die Pupillenreaktion von Probanden während der Bearbeitung verschiedener Aufgaben (z. B. Multiplikation) mit einem Remote Eye Tracker. Die deskriptive Ergebnisdarstellung zeigt zwar Unterschiede in der Pupillenreaktion in Abhängigkeit verschiedener Schwierigkeitsgrade, allerdings betonen die Autoren die kritische Validität der Daten aufgrund oben genannter methodischer Probleme, insbesondere bezüglich des Einflusses durch Beleuchtung. Entsprechend scheint sich Eye Tracking als Methode zur Aufzeichnung der Pupillenreaktion weniger gut zu eignen (Poole & Ball, 2000). Darüber hinaus gilt zu beachten, dass die Pupillenreaktion als kognitiver Indikator durch das autonome Nervensystem gesteuert wird und damit keiner bewussten Kontrolle unterliegt. Im Gegensatz dazu können Scanning Techniken von Personen intendiert gelenkt und manipuliert werden. Damit unterscheiden sich die Ansätze der Pupillometrie und des Eye Trackings per se. Die Güte der Pupillometrie als Online Messmethode für kognitive Prozesse ist zusammenfassend eher kritisch zu bewerten.

Eine weitere Methode zur Aufzeichnung von Augenbewegungen ist die **Elektrookulografie** mittels Elektrookulografen (EOG; Schandry, 1989, S. 276; Merchant,

2001), welche auf der Tatsache basiert, dass zwischen Horn- und Netzhaut eine ableitbare Potentialdifferenz existiert. Problematisch an dieser Methode zur Registrierung der Blicke auf ein bestimmtes Objekt ist, dass lediglich indirekt von der Blickrichtung der Probanden und entsprechenden Ausschlägen des EOGs auf fixierte Punkte geschlossen werden kann. Damit ist die Datenauswertung mit Subjektivität kontaminiert, die bei der Eye Tracking Methode aufgrund zweier überlagerter Videobilder (Feldkamera und Augenkamera) und Koordinatenberechnungen an dieser Stelle geringer sein dürfte. Laut Schandry (1989, S. 281ff.) ist der über EOG gemessene Lidschlag ein „*Korrelat der Schreckreaktion*“ und „*kann als Aktivierungsindikator dienen*“. Interne kognitive Prozesse, wie z. B. Kopfrechnen oder Emotionen können zu einer Erhöhung der Lidschlagfrequenz führen, während externe Stimuli (z. B. visuelle Suchaufgaben) zu einer Verminderung führen. Zu beachten ist, dass die aufgabenunabhängige Lidschlagfrequenz interindividuell stark variiert und auch durch z. B. Verschmutzungen des Auges beeinflusst sein kann. Entsprechend fassen Poole und Ball (2000, S. 7) zusammen: „*For these reasons, pupil size and blink rate are less often used in eye tracking research*“. Bei Tsai et al. (2007) findet sich eine Übersicht über empirische Studien, die den Zusammenhang der Lidschlagfrequenz mit Leistung analysieren, wobei die Befunde erheblich inkonsistent sind. Die Lidschlagfrequenz als Indikator für mentale Beanspruchung bedarf also ähnlich wie die Pupillendilatation weiterer Forschungsbemühungen.

Es existieren daneben eher seltene Methoden, mit denen Blickbewegungsparameter aufgezeichnet werden können. Hierunter zählen z. B. **Limbus Tracking** (Detektion der Bewegungen des Limbus²³) oder Tracking via **Kontaktlinsen**, die mit Lichtreflexzonen ausgestattet sind, sodass von Lichtpunkten auf dem Auge auf die Blickrichtung geschlossen werden kann. Insbesondere bei letztgenannter Methode dürfte die Akzeptanz durch Versuchsteilnehmer eher gering im Vergleich zu „herkömmlichem“ Eye Tracking sein. Auf die Beispiele soll hier aufgrund ihrer seltenen Anwendung nicht weiter eingegangen werden; näher erklärt werden sie bei Merchant (2001). Für die vorliegende Arbeit, die sich insbesondere auf Scanning Techniken von Hubschrauberpiloten bezieht, ist ohnehin ausschließlich das Eye Tracking Verfahren indiziert.

²³ Unter dem Limbus wird der Übergangsbereich verstanden, der sich zwischen Cornea und Sclera (also der Lederhaut) des Augapfels befindet.

2.1.8 Anwendungsbereiche von Eye Tracking

Um neben der Anwendung von Eye Tracking in Flugsimulationsumgebungen eine erweiterte Perspektive auf das Verfahren zu ermöglichen, werden die Kernfachgebiete, in denen es zum Einsatz kommt, exemplarisch vorgestellt.

Im Rahmen der **Sicherheitsevaluation von Mensch-Maschine-Systemen** spielt Eye Tracking im Bereich der Flugsicherung (Air Traffic Control) oder in der Automobilindustrie (z. B. Fahrerassistenzsysteme; FAS) eine immer größere Rolle. Hintergrund der wachsenden Anwendungshäufigkeit ist die steigende Automation interaktiver Systeme (Duchowski, 2002), die immer weniger aktive Handlung durch den Nutzer fordern, sondern vielmehr dessen visuelle Aufmerksamkeit und ein Regulieren im richtigen Moment (Merchant, 2001). Von Interesse ist daher, wohin der Blick des Nutzers unter bestimmten Anzeige-Bedingungen und zu einem bestimmten Zeitpunkt gerichtet ist und inwiefern dadurch seine Aufmerksamkeit und Entscheidungen beeinflusst werden. Entsprechend hat Eye Tracking hier ein hohes bildgebendes Potenzial zum Verständnis sicherheitskritischer kognitiver Zustände.

Beziehungen zwischen Kognition und Scanning Techniken werden teilweise mithilfe **kognitiver Modellierung**, die mentale Prozesse simulieren und mathematisch vorhersagen, analysiert. Die beiden Modelle ACT-R (Adaptive Control of Thought-Rational) und HMM (Hidden Markov Models) werden dabei beispielsweise herangezogen. Im empirischen Vergleich einer ACT-R Verhaltensvorhersage mit Eye Tracking Daten bei einer visuellen Suchaufgabe zeigt sich eine mittlere Vorhersageleistung durch das Modell (52 %; s. Byrne, 2001 und vgl. Salvucci, 2000). Im Gegensatz zu ACT-R Modellen, die über Lernprozesse Wissen modellieren, handelt es sich bei HMMs um dynamische, stochastische Modelle, die sich durch zwei Zufallsprozesse beschreiben lassen und die versuchen, Fixationen oder Blicke in Vektorräumen zu identifizieren. Die HMM Analyse der Scanning Techniken von Piloten in einer Studie von Hayashi (2003) erbringt eine Übereinstimmung von bis zu 89 % mit den empirischen Eye Tracking Daten. Hayashi, Oman und Zuschlag (2004) zeigen die Anwendbarkeit der HMM Methode außerdem in einer HUD-Evaluationsstudie. Die Weiterentwicklung von kognitiven Vorhersagemodellen bietet eine potenzielle Möglichkeit dafür, Blickalgorithmen im Mensch-Maschine-Arbeitsplatz zu implementieren, um damit ein adäquates Beanspruchungsniveau für Operateure zu schaffen. Darüber hinaus kann mit solchen Modellen der Weg hin zur Blicksteuerung von technischen Systemen ausgebaut werden.

Weitere Anwendungsbereiche von Eye Tracking sind beispielsweise die Evaluation der Wirkung von (Print-)Werbung durch die Analyse der Scanning Techniken in der **Marktforschung** oder die **Gebrauchstauglichkeitsbewertung** von Produkten,

Software und Webseiten (Duchowski, 2002; Geise, 2012). Innerhalb dieser Einsatzfelder werden zumeist mehrere Varianten von Produkten oder beispielsweise Webseiten hinsichtlich der Scanning Techniken von Testpersonen miteinander verglichen. Ein Testgegenstand gilt dann als interessanter, wirksamer oder verkaufsträchtiger, wenn Testpersonen ihn länger und häufiger betrachten als einen Referenztestgegenstand.

Als nützlich erwiesen hat sich Eye Tracking außerdem in der **Leseforschung**; einen ausführlichen Überblick liefert Rayner (1985). Es kann gezeigt werden, dass sich die Scanning Techniken beim stillen und lauten Lesen unterscheiden: Beim lauten Lesen oder Mitlesen während eine andere Person vorliest, ist die Fixationsdauer länger als beim stillen und die Augen tendieren dazu, schneller zu lesen als Akteure das Gelesene verbalisieren können. Im Rahmen des stillen Lesens dauern Fixationen etwa 200 – 250 ms, mit Sakkaden von etwa 7 – 9 Buchstaben. Während die meisten Wörter fixiert werden, lassen Leser auch viele aus, sodass nicht immer foveale Fixation notwendig ist (Funktionswörter werden zu 35 % und damit weniger häufig fixiert, Inhaltswörter hingegen zu 85 %). Bezüglich der Sakkaden wurde empirisch beobachtet, dass 10 – 15 % der Sakkaden beim Lesen rückläufig (regressiv) sind. Innerhalb eines Wortes wird zurückgesprungen, wenn die vollständige Verarbeitung fehlschlug. Beim Wechsel in die nächste Zeile müssen häufig korrigierende Bewegungen vollzogen werden, wobei eine Fixation am Zeilenbeginn nicht direkt, sondern immer 5 – 7 Buchstaben entfernt stattfindet. Wenn ein Text inhaltlich schwieriger wird, steigt die Fixationsdauer, die Sakkadenlänge nimmt ab und die Frequenz der Regressionen nimmt zu. Eye Tracking beim Lesen ermöglicht außerdem die Differenzierung von guten und schlechten Lesern bzw. die Detektion von Dyslexie: Schlechte, ungeübte und dyslektische Leser zeigen längere und mehr Fixationen, kürzere Sakkaden und mehr Regressionen als normale Leser. Hier beweist das Verfahren also diagnostische Güte.

Darüber hinaus bildet Eye Tracking die Basis für die **visuelle Steuerung** von Assistenzsystemen durch behinderte Menschen (z. B. Poole & Ball, 2000; DeSantis et al., 2005). Ein noch kritischer Punkt ist hierbei, dass die Blicksteuerung bei geübten Nutzern zwar eine schnellere Bedienung eines Systems ermöglichen kann, der Blickort von Ungeübten aber häufig nicht so präzise lokalisiert wird wie ein Mausklick (Duchowski, 2002). Am letztgenannten Beispiel wird deutlich, dass Eye Tracking nicht nur genutzt wird, um Blickdaten zu erheben, sondern, wie in Kapitel 1 schon dargestellt, „as an input device“ dienlich ist (Poole & Ball, 2000, S. 8).

2.1.9 Zusammenfassung technischer Aspekte

In Korrespondenz zu den multiplen Ausprägungen von Eye Tracking Systemen und den damit zu erfassenden Parametern prosperieren die Einsatzbereiche und methodischen Diskussionen darum. Wünschenswert ist in mittelfristiger Zukunft eine detaillierte Standardisierung in Form von klar definierten Schwellenbereichen für die Ausprägung von Blickparametern. Während sich Eye Tracking als sensibles, bildgebendes Werkzeug bereits mehrfach beweisen konnte, bedarf es hinsichtlich Ökonomie und Fehlertoleranz der Messeinrichtungen bei realistischen Versuchsumgebungen (z. B. bei Situationen mit viel Kopfbewegung der Probanden) weiterer technischer Entwicklungsarbeit.

2.2 Theoretische und empirische Grundlagen zu Eye Tracking

Um für diese Arbeit einen theoretischen Rahmen zu schaffen, werden zunächst Grundlagen der menschlichen Informationsverarbeitung betrachtet, da diese bei der Flugführung den basalen Nukleus des (Entscheidungs-) Verhaltens bilden. Im Vordergrund stehen im Folgenden Informationsverarbeitungsmodelle, wobei Konstrukte wie (visuelle) Aufmerksamkeit, Wahrnehmung, Arbeitsgedächtnis und Aufgabenanforderungen sowie Langzeitgedächtnis und Erfahrung zentrale Einflussfaktoren bilden, die hier vorgestellt werden. Diskutiert wird auch der Einfluss bisher eher unbeachteter Konzepte auf die Informationsverarbeitung, wie Emotionen, Motivation und Ermüdung. Bisherige Befunde aus Eye Tracking Studien werden in die theoretische Abhandlung integriert um daraus Hypothesen über Scanning Techniken von Hubschrauberpiloten ableiten zu können. Da nicht für jedes theoretische Konzept Studien aus der Luftfahrt vorliegen, werden auch Erkenntnisse aus der Automobilbranche einbezogen. Hinführend wird zunächst ein Überblick über die Aufgabe der Flugführung geliefert.

2.2.1 Flugführung als anspruchsvolle Prozessführung

Die Flugführung als Mensch-Maschine-Interaktion kann als Prozessführung durch den Operateur, bei der die „Überwachung und Steuerung von dynamischen Prozessen“ (Herczeg, 2005, S. 88; vgl. Rasmussen, 1986) im Vordergrund steht, betrachtet werden. Die Anforderungen, die dabei für Piloten entstehen, werden anhand des folgenden Exkurses verdeutlicht.

Was versteht man unter einer Mission?

Eine typische Flugaufgabe für Kampfpiloten ist die Luftraumverteidigung im Sichtflug (VFR; Svensson, Angelborg-Thanderz & Wilson, 1999). Denning, Bennett und Crane (2003, S. 3) spezifizieren einige Zielstellungen solcher Missionen: *“Typically,*

the objective of real-world missions involving air combat will be to establish air superiority, block incursions into an opening area, and deny the use of selected airspace for some period of time. The degree of aggressiveness of adversaries, combined with logistical constraints of weapons and fuel, and psychological pressures on both sides will temper the degree of confrontation.“ Ganz andere Herausforderungen treten beispielsweise bei Missionen von Polizei-Hubschrauberpiloten auf: *“[Ich erhielt den] Auftrag, eine herzkrankte Person zu evakuieren. Dieser Patient befand sich auf einem 60 m langen Fischerboot [bei einer] Sicht von 1 NM [...]. Am Einsatzort war es meine Absicht ein OWTD (overwater transition down) zum Boot durchzuführen, um es mit dem bordeigenen Radar zu identifizieren. [...] Wir wollten dann den Patienten mit der Winde aufnehmen, IFR [mit Instrumentenflug] zur Betankung nach Naine fliegen und im Anschluss zurück nach Goose Bay, wo er in ein Krankenhaus eingeliefert werden sollte.“* (Noble, 2012, Magazin des Bundespolizei-Flugdienstes, S. 40f). Die Belastungen für fliegerisches Personal bei der Umsetzung solcher Missionen differenziert Buld (2000, S. 90ff) in verschiedene Dimensionen, von denen einige lauten:

- *„Entscheiden müssen“ („Bei der Bekämpfung von Zielen [...]: Bin ich überhaupt bedroht? Jetzt schießen oder warten, bis das Ziel noch besser getroffen werden kann?“),*
- *„Mehrfach Tätigkeiten ausführen“ („Hubschrauber – Gebirgsflug: Bordmechaniker muss Winde bedienen, Lage des Hubschraubers im Blick behalten, Pilot lotsen“),*
- *„in Zusammenhängen denken“ („Luftkampf: Wie ist die eigene Lage im Raum? Wie viele Flugzeuge sind noch in diesem Luftraum? Welche davon sind Feind? Wo befinden sich diese? Was tut der Feind, wenn ich XY tue?“),*
- *„Aufgaben präzise ausführen“ („Einweisen des Flugzeugführers durch den Bordmechaniker bei SAR [Safe and Rescue] im Gebirge, Präzision bei Anflugmanövern“),*
- *„Körperlichen Anforderungen standhalten“ („Vibrationen“, „ungünstige Klimabedingungen“),*
- *„Monotonie ertragen“ („Transportflüge: 6 Stunden fliegen bei gleichförmigem Geräuschpegel und trotzdem präzisen Anflug durchführen“).*

2.2.2 Ein Grundmodell der Informationsverarbeitung

Einen psychologischen Kernprozess, der der Erfüllung der oben genannten Anforderungen von fliegerischen Tätigkeiten vorausgeht, bildet die menschliche Informationsverarbeitung, die ihrerseits der Analyse von Scanning Techniken zugrunde

liegt. Informationsverarbeitung wird häufig als sequenzieller Prozess betrachtet, dessen Stufen sich in Phasenmodellen modellieren lassen. In der Luftfahrt z. B. wird der Pilot traditionell aus der Perspektive eines Informationsverarbeitungssystems betrachtet; in den Worten von Wickens (1986, zitiert nach ebd., 2009) wird er auch als ein „**information processor**“ charakterisiert (s. Abbildung 13).

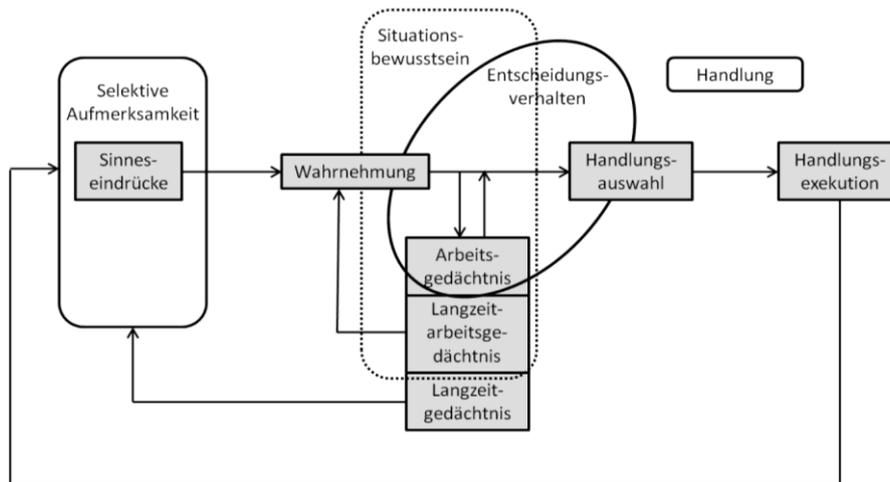


Abbildung 13: Grundmodell der Informationsverarbeitung (übersetzt aus Wickens, 2009).

Wickens (ebd.) sieht **selektive Aufmerksamkeit** innerhalb der einzelnen Sinnesmodalitäten als Voraussetzung zur Akquise einer begrenzten Anzahl an Informationen. Diese Informationen werden unter Zuhilfenahme von **Erfahrungen** aus dem **Langzeitgedächtnis** und dem **Langzeitarbeitsgedächtnis**²⁴ wahrgenommen, interpretiert und verstanden. Die Wahrnehmung und Verarbeitung von Informationen kann zur **Auswahl und Ausführung einer Handlung** führen (z. B. Korrektur der Flughöhe nach Feststellung von Höhenverlust), hat aber nicht in jedem Fall eine Aktion des Operators zur Folge. Vielmehr ist es auch möglich, dass Informationen eine bestimmte Zeit lang im **Arbeitsgedächtnis** gehalten (z. B. eine Anweisung des Towers) oder im Langzeitgedächtnis abgespeichert werden (z. B. ein vorausgehender technischer Zwischenfall mit dem Luftfahrzeug). Viele Sinneseindrücke gehen aber aufgrund der Begrenztheit der Aufmerksamkeit ohne weitere Verarbeitung wieder verloren.

Das Phasenmodell von Wickens (ebd.) unterstellt sowohl eine sequenzielle Abfolge von der Aufmerksamkeit bis hin zur Handlungsexekution als auch einen **Rückkopplungsmechanismus**, wonach Informationsverarbeitung zwar ein kontinuierli-

²⁴ Unter dem Langzeitarbeitsgedächtnis wird der Teil des Arbeitsgedächtnisses verstanden, der aufgrund von Erfahrung effektiv auf aufgabenspezifische Inhalte aus dem Langzeitgedächtnis zurückgreifen kann (z. B. Schachexperten; Ericsson & Kintsch, 1995).

cher Prozess ist, aber auf jeder Stufe des Modells Aktivitäten des Operateurs ausgelöst werden können (z. B. spontanes Überprüfen der Flugroute im Navigationssystem). Außerdem berücksichtigt das aktuelle Modell von Wickens (2009) das Konstrukt des **Situationsbewusstseins** (Endsley, z. B. 1995; 1996; 2000), worunter die korrekte Wahrnehmung, Interpretation und Vorhersage einer Situation verstanden wird. Die Ausprägung des Situationsbewusstseins beeinflusst maßgeblich die **Entscheidungen** des Piloten (Aeronautical Decision Making) sowie die Auswahl und Exekution einer Handlung. Wickens betont außerdem den Einfluss von **Aufgabeanforderungen** durch Missionsinhalte oder Automation auf die Informationsverarbeitung. Aufgrund begrenzter Kapazitäten des Arbeitsgedächtnisses können diese zur subjektiven **Beanspruchung (Workload)** des Piloten beitragen. Wickens führt beispielsweise an, dass sich Piloten, die über Start oder Abbruch wegen technischer Auffälligkeiten im Cockpit entscheiden müssen, in einer Situation mit hoher Workload befinden.

Zur Beschreibung, Erklärung und Vorhersage von Verhalten im Arbeitskontext existieren neben typischen Informationsverarbeitungsmodellen wie dem von Wickens (1986, 2009) **Handlungsregulationstheorien**, die Handlungen entweder als phasisch modellieren oder strukturell zerlegen (z. B. Hacker, 1973; Volpert, 1974; Rasmussen, 1986). Die Modelle beschreiben die Handlungsregulation besonders in der sicherheitskritischen Prozessführung (High Risk Environments), wie z. B. in Kernkraftwerken, in der Medizin oder in der Luftfahrt. Die teilweise sehr differenzierten und umfangreichen Theorien sollen hier aus Relevanzgründen nicht im Einzelnen dargestellt werden, da hier die Ebenen der visuellen Aufmerksamkeit und Wahrnehmung im Vordergrund stehen. Hingewiesen sei darauf, dass die Modelle komplexe Arbeitshandlungen als in Einheiten zerlegbar begreifen und diese jeweils über Rückkopplungsprozesse gesteuert werden (z. B. Vergleichs-Veränderungs-Rückkoppelungs-[VVR]-Einheiten bei Hacker, 1973). Einbezogen werden sowohl automatisierte Verhaltensweisen auf sensomotorischer Regulationsebene (z. B. Skill-Based Behavior bei Rasmussen, 1986) als auch wissensbasiertes Verhalten als planvolle und zielgerichtete Handlungen (z. B. Knowledge-Based Behavior bei Rasmussen, 1986). Das Prinzip der Rückkopplung von Handlungen auf die Informationsaufnahme und -verarbeitung findet sich analog bei Wickens (2009, s. oben) in Form einer Feedbackschleife.

Bevor detailliert auf die von Wickens genannten Komponenten der Informationsverarbeitung (Aufmerksamkeit und Wahrnehmung, Arbeitsgedächtnis, Aufgabenanforderungen, Langzeitgedächtnis und Erfahrung sowie Situationsbewusstsein) und deren Rolle bei der Modulation von Scanning Techniken eingegangen werden soll, wird zunächst beleuchtet, ob die erwähnten Modelle den Prozess der Informationsverarbeitung und Handlungsregulation erschöpfend beschreiben.

2.2.3 Ergänzungen zum Modell des „Information Processors“

Folgt man Herczeg und Stein (2012), ist den meisten Informationsverarbeitungs- und Handlungsregulationsmodellen gemein, dass sie den Mensch aus einer eher mechanistischen Perspektive betrachten, der, in Analogie zu einem Computer, Informationen hierarchisch und rational verarbeitet. Die Autoren (ebd.) weisen darauf hin, dass der Mensch dabei aber nicht nur als rationaler Denker betrachtet werden kann, sondern auch Einflüsse wie Emotionen, Motivation, Ressourcenmanagement (Effort Regulation, Workload), Ermüdung und Stress sowie soziale und organisationale Prozesse berücksichtigt werden müssen, um menschliches Handeln zu beschreiben. Die genannten Einflüsse können theoretisch auf allen Ebenen der Informationsverarbeitung wirksam werden und dazu führen, dass menschliches Verhalten im Gegensatz zu einem Computer nur schwer vorhersagbar wird. Herczeg und Stein (2012) zufolge wird Informationsverarbeitung in den klassischen Ansätzen zu stark als eindimensionaler Prozess beschrieben. Es sei aber vielmehr so, dass Operateure auch zwischen den Stufen der Informationsverarbeitung hin- und herspringen und Verarbeitungsprozesse verkürzen können (Rule-Based Shortcuts nach Rasmussen, 1986). Außerdem ließen sich Handlungen nicht immer strukturiert zerlegen, sondern Leistungsergebnisse müssten aufgrund mentaler Modelle erfahrener Operateure eher holistisch betrachtet werden. Die Autoren plädieren für die Kombination kognitiver und emotionaler Modelle bei der Beschreibung von Operateurverhalten und sprechen sich für eine komplexere theoretische Vorstellung²⁵ menschlichen Handelns aus. Der integrative Ansatz von Herczeg und Stein (ebd.) bleibt auf theoretischer Ebene und bildet eher allgemeine Hypothesen über beispielsweise bidirektionale Einflüsse sowie Rückkopplungsschleifen von Emotionen auf die Interaktion zwischen Mensch und Maschine. Inwiefern sich für den Ansatz im Kontext der Luftfahrzeugführung **empirische Hinweise** finden lassen, wird im Folgenden geprüft. Dabei steht der Einfluss von Emotionen, Motivation sowie Ermüdung auf der Ebene von Aufmerksamkeit und Wahrnehmung im Fokus des Interesses. Es gilt konkret herauszufinden, ob affektive, motivationale und ermüdungsbezogene Komponenten tatsächlich die Scanning Techniken von Hubschrauberpiloten beeinflussen können.

²⁵ Dem Anspruch auf Vollständigkeit halber sollen auch externe Einflüsse auf die Informationsverarbeitung von Piloten Erwähnung finden; hier unter anderem das Cockpitdesign, die Wirkung psychotroper Substanzen (Koffein, Alkohol), Umgebungsbedingungen (Wetter, Zusammensetzung der Luft, Dunkelheit) oder G-Kräfte. An den Beispielen wird deutlich, dass ein Modell zur Beschreibung menschlicher Informationsverarbeitung kaum erschöpfend sein kann/wird.

2.2.4 Einfluss affektiver und motivationaler Komponenten

In den Modellen zur Beschreibung von Pilotenverhalten hatten **Emotionen** bisher einen eher marginalen Status. Im ausführlichen Artikel von Hudlicka und McNeese (2002) wird zunächst allgemein darauf hingewiesen, dass affektive Zustände perzeptuelle, kognitive und motorische Prozesse beeinflussen können. Die Autoren zählen exemplarische Effekte auf:

- *“Altering the nature of attentional processing (e.g. change in focal area, increased/reduced size of focal area, biasing attention towards or away from particular stimuli, etc.);*
- *Helping to activate (or inhibit) particular perceptual and cognitive schemata that enhance (or limit) the perception or processing of specific stimuli;*
- *Promoting (or inhibiting) the selection of particular actions, and influencing the accuracy and speed of selected motor responses”* (Hudlicka & McNeese, ebd, S. 2).

Hudlicka und McNeese (2002) zeigen anhand einer Reihe von Befunden die Gültigkeit ihrer Beispiele (z. B. Einfluss von Angst auf die Aufmerksamkeit, von Emotionen auf Gedächtnis, von Zwanghaftigkeit auf die Leistung und von Heuristiken auf Entscheidungsverhalten). Allerdings stammen hierbei viele Studien aus Tierversuchen, aus dem Labor-Kontext oder wurden mit psychisch Erkrankten durchgeführt (z. B. Kimble, Fleming, Brandy, Kim, & Zambetti, 2010; Shaw & Porter, 2012; Bayram & Esgin, 2013). Daher ist schwer abzuleiten, ob und inwiefern Emotionen konkret die Scanning Techniken von Hubschrauberpiloten beeinflussen können. Hudlicka und McNeese (ebd.) führen zwar hypothetische Szenarien an, in denen (Kampf-)Piloten aus Angst eine mehrdeutige Radarinformation fehlinterpretieren oder Friendly Fire durchführen können, liefern aber keinen empirischen Nachweis hierfür. Es ist bei der Erforschung solcher durch Emotionen ausgelösten Verzerrungen zu beachten, dass affektive Stati bei den Probanden durch geschickte Manipulation induziert (z. B. durch Zwischenfälle bei den Missionen oder die Simulation von Beschuss) und kontrolliert (z. B. durch Interview oder physiologische Messungen) werden müssen. Dies würde die Fragestellung der vorliegenden Studie übersteigen, da hier die Scanning Techniken von Hubschrauberpiloten bei regulären Landemanövern im Vordergrund stehen. Für Folgeuntersuchungen wäre der Einfluss von Emotionen auf Scanning Techniken im Hubschraubercockpit sicherlich eine interessante Fragestellung. Methodisch zu bedenken ist aber außerdem, dass Scanning Techniken nicht unbedingt ein geeigneter Indikator für affektive Reaktionen sind: Rohani (2012) fand z. B., dass Angst zwar Aufmerksamkeitsfehler bedingen kann (Kinder erhalten zur Angstinduktion Bedrohungsinformationen über Tiere und finden im Anschluss weniger Zielreize), aber hier war die

Reaktionszeit aussagekräftiger als Eye Tracking Daten. Trotzdem konnte bereits nachgewiesen werden, dass emotionale Bilder im Vergleich zu neutralen eine größere Anzahl an Fixationen und eine größere Sakkadenlänge erzeugen können (Bradley, Keil, & Lang, 2012). Neben diesem Einfluss von Emotionen auf die Scanning Techniken beim Betrachten von affektivem Material im Alltag muss aber bedacht werden, dass diese Situationen nicht uneingeschränkt auf die Flugführung übertragen werden können. Bei einer Auswertung des NASA-Sicherheitsmeldesystems hinsichtlich subjektiver Störeinflüsse auf das Situationsbewusstsein von Piloten fanden Gibson, Orasanu, Villeda und Nygren (1997), dass Emotionen nur in etwa 7 % der Zwischenfälle in der Zivilluftfahrt eine Rolle spielten. Hauptstörfaktoren waren hier eher Workload (86 %), Kommunikation und Koordination (74 %) sowie unangemessene Verfahren (54 %). Sicherlich könnten diese Angaben sozialen Erwünschtheitstendenzen unterliegen, aber gegen den Einfluss von Emotionen auf die Scanning Techniken beim Wahrnehmen gewöhnlicher Landschaften sprechen insbesondere auch die Befunde von Pannasch, Helmert, Roth, Herbold und Walter (2008). Sie untersuchten den Einfluss von positiver oder negativer Stimmung auf die Scanning Technik beim Betrachten von Landschaftsbildern. Über ein standardisiertes Verfahren (Bericht eines positiven bzw. negativen Lebensereignisses sowie musikalische Induktion) wurde bei Probanden positive oder negative Stimmung erzeugt und über Fragebogen kontrolliert. Die Scanning Techniken der Personen wurden über Eye Tracking erfasst, Sakkaden und Fixationsdauer in einem Prä-/Posttestvergleich überprüft. Es zeigte sich (S. 13): *„In conclusion, the manipulation of mood, regardless if positive or negative, has **no substantial influence** on the characteristics of both gaze parameters in early and late processing phases.“*

Motivationale Einflüsse können die Verarbeitung und Interpretation visueller Reize beeinflussen. Menschen, die durstig sind, sehen in einem mehrdeutigen Stimulus beispielsweise mehr Transparenz (eine Eigenschaft, die mit Wasser assoziierbar ist) als Menschen, die nicht durstig sind (Changizi & Hall, 2001). Balceitis und Dunning (2006) nennen eine Reihe weiterer Studienergebnisse, wonach Probanden nach einem motivationalen Priming (Induktion eines Bedürfnisses, z. B. Durst oder sozialer Anschluss) ambivalente Situationen stets so wahrnehmen, wie sie sie sich wünschen. Dies untermauert, dass motivationale Tendenzen auf der Stufe der Interpretation von Reizen die Informationsverarbeitung verzerren können. Ob motivationale Prozesse tatsächlich auch a) auf der Stufe der visuellen Aufmerksamkeit und Informationsakquise wirksam werden und b) damit die Scanning Techniken von Hubschrauberpiloten beeinflussen können, **bleibt derzeit unbeantwortet**. Auch hier müsste, um spezielle Ergebnisse zu erzielen, eine Motivationsinduktion und Kontrolle stattfinden (z. B. Bedürfnis einen verunglückten

Kameraden zu bergen) und eine mehrdeutige Landschaft beim Hubschrauberflug geschaffen werden (z. B. Felsbrocken, die aussehen wie ein regloser, menschlicher Körper). So könnte verglichen werden, ob Personen mit Motivationsinduktion den Felsbrocken länger fixieren als Piloten ohne Rettungsbedürfnis. Ein solches Untersuchungsdesign erlaubte es aber ebenfalls nicht, Scanning Techniken von Hubschrauberpiloten zu erforschen, die zunächst nur der Motivation folgen, einen Hubschrauber regulär zu landen.

Neben Einflüssen wie Emotion und Motivation können auch **Ermüdung** und Stress eine Wirkung auf die Informationsverarbeitung ausüben: „*Fatigue degrades not only basic cognitive performance [...] but also flight performance, including the ability to maintain designated flight parameters [...]*“ (Previc, Lopez, Ercoline, Daluz, Workman, Evans, & Dillon, 2009, S. 326f). Während Ermüdung also zu einem Abfall der Gesamtleistung bei einem Flug führen kann, lässt sich bisher nicht nachweisen, dass die Scanning Techniken von Piloten hiervon betroffen sind. Previc et al. (2009) finden nämlich ein paradoxes Ergebnis: Während die Scanning Techniken nach 30 Stunden Wachheit mit zehn aufeinanderfolgenden Simulatorflügen (N = 10 Zivilpiloten) nicht beeinträchtigt sind, berichten die Probanden von einem starken Anstieg der subjektiven Ermüdung. Dies lässt darauf schließen, dass Piloten in der Lage sind, trotz wahrgenommenen Ermüdungserscheinungen ihre bewährten Werkzeuge (hier: Scanning Techniken) zur Informationsakquise einzusetzen um damit den Flug zu bewältigen. Bemerkenswert war in der Studie die Tatsache, dass die subjektive Ermüdung erst nach sechs Stunden signifikant anstieg. Die Autoren fassen zusammen: „[...] *we did not observe changes in any basic eye-movement parameter even though our average values were comparable to others*“ (Previc et al., 2009, S. 343). Insofern lässt sich der Einfluss von Ermüdung auf die Leistung nicht leugnen, aber eine Beeinflussung der Scanning Techniken wird empirisch nicht angezeigt, da offenbar gilt: „*instrument scanning [...] is a highly practiced behavior that is **resistant to fatigue***“ (ebd, S. 344).

Zusammengefasst lässt sich festhalten, dass klassische Informationsverarbeitungs- und Handlungsregulationsmodelle den Einfluss von affektiven und motivationalen Komponenten bei der Erklärung menschlichen Verhaltens unterschlagen. Bei der Interpretation von Stimuli, dem Treffen von Handlungsentscheidungen sowie der Handlungsexekution ist die Unterstellung emotionaler und motivationaler Effekte **empirisch haltbar**. Auf der Ebene der visuellen Aufmerksamkeit und insbesondere in Bezug auf den Einsatz von Scanning Techniken zur Informationsakquise kann der Einfluss dieser Faktoren auf Basis recherchierter Befunde hingegen **nicht bestätigt** werden. Zur Analyse der Scanning Techniken von Hubschrauberführern bei regulären Landemanövern werden affektive, motivations- und ermüdungsbezoge-

ne Einflüsse daher nicht ausführlicher verfolgt. Vielmehr stehen im Folgenden die aufmerksamkeits-, wahrnehmungs- und gedächtnisbezogenen Elemente aus dem Grundlagenmodell zur Informationsverarbeitung (Wickens, 2009) im Vordergrund.

2.2.5 Aufmerksamkeit und Wahrnehmung

Aufmerksamkeit, sensorische Verarbeitung und Wahrnehmung von Informationen stehen am Beginn des Modells von Wickens (2009, s. Abbildung 14).

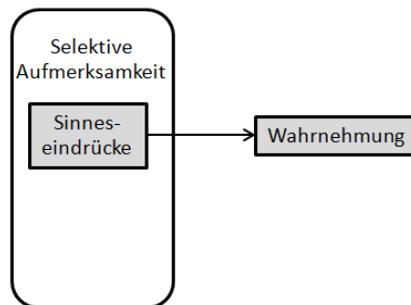


Abbildung 14: Navigation im Modell von Wickens (2009): Aufmerksamkeit und Wahrnehmung.

Nach Anderson (1989, S. 51) determiniert Aufmerksamkeit die Informationsverarbeitung: Große Mengen an Informationen können zwar in den sensorischen Speicher gelangen, wenn jedoch keine Aufmerksamkeit auf sie gerichtet ist, gehen sie wieder verloren. Mit anderen Worten: „*Attention is the preferential processing of some items to the detriment of others*“ (Findlay & Gilchrist, 2003, S. 3; zitiert nach Geise 2012, S. 9). Geise deutet hier ein Paradox an, wonach Aufmerksamkeit auf der einen Seite selektiv und begrenzt ist, auf der anderen aber mit einer zeitweisen Aktivierungserhöhung einhergeht, bei der Menschen gegenüber bestimmten Stimuli sensibilisiert sind. Hinsichtlich der Selektivität der Aufmerksamkeit herrscht theoretische Heterogenität im Hinblick auf die Ebene der Selektion (vgl. Broadbent, 1958; Treisman, 1960; Deutsch & Deutsch, 1963). Bak (1999) hält fest, dass die Auswahl früh nach der Reizaufnahme oder auch erst später nach semantischer Enkodierung stattfinden kann (Early vs. Late Selection). Spätere Selektion erfordere dabei mehr Ressourcen als frühere: „*Es kann angenommen werden, daß die Selektion bei ausreichenden Verarbeitungsressourcen spät erfolgt, während bei reduzierten Ressourcen früh selektiert wird [...]*“ (Bak, 1999, S. 4).

Zur Begrenztheit der Aufmerksamkeit existieren verschiedene theoretische Perspektiven, die sich in zwei größere Strömungen unterteilen lassen: Auf der einen Seite wird Aufmerksamkeit in frühen Filtermodellen „*[...] als eine sehr begrenzte Ressource verstanden [...], [die] auf genau einen Inhalt beschränkt [ist]*“ (Anderson, 1989, S. 52). Auf der anderen Seite existieren Ansätze, die zwar auch von begrenzten Ressourcen ausgehen, aber gleichzeitig postulieren, dass mehrere

Aufgaben simultan ausgeführt werden können, wobei diese Fähigkeit dadurch determiniert wird, wie viel Kapazität die Aufgabe beansprucht (Kahneman, 1973).

In der **Filtertheorie der Aufmerksamkeit** von Broadbent (1958) wird davon ausgegangen, dass gleichzeitig dargebotene Sinnesreize simultan in den sensorischen Speicher gelangen, wo sie aufgrund ihrer Eigenschaften (z. B. Stimmlage), nicht aber ihrer semantischen Bedeutung, früh selektiert und entsprechend weiter verarbeitet werden. Die Filter- oder auch Flaschenhalstheorie nimmt an, dass alle anderen Reize aus dem sensorischen Speicher verloren gehen (s. Abbildung 15).



Abbildung 15: Flaschenhalstheorie²⁶.

In Bezug auf die fliegerische Tätigkeit bedeutete dies beispielsweise, dass Piloten, die sich auf ein Cockpit-Display konzentrieren, nicht in der Lage wären, die Funkanweisungen des Towers zu verarbeiten. Gegen diese eher eindimensionale Perspektive bezüglich der Aufmerksamkeit sprechen die früh erforschten Phänomene der **selektiven** (dichotisches Hören) oder **geteilten Aufmerksamkeit** (Cocktailparty-Phänomen). Bei Experimenten zum dichotischen Hören (z. B. Grey & Wedderburn, 1960; Treisman, 1960) wurden Probanden gebeten, einen von zwei verschiedenen Texten, die jeweils auf einem Ohr zu hören sind, zu beschatten. Gezeigt wurde, dass Personen ihre Aufmerksamkeit selektiv zwischen den beiden Nachrichten hin und her richten können: Während der eine Text korrekt wiedergegeben wird, kann die andere Botschaft vollständig ausgeblendet werden. Menschen sind außerdem in der Lage ihre Aufmerksamkeit zwischen zwei Gesprächen (z. B. inmitten einer Cocktailparty; Cherry, 1953) aufzuteilen: Fällt in einem entfernten Gespräch der eigene Name (oder ein anderer bedeutsamer Stimulus) während man sich gerade selbst in einem Gespräch befindet, wechselt die Aufmerksamkeit kurzfristig zu dem entfernten Gespräch, obwohl sie auf das eigene Gespräch gerichtet war. Diesen Erkenntnissen zufolge entwickelte Treisman (1960) die **Dämpfungstheorie der Aufmerksamkeit**, die davon ausgeht, dass scheinbar unbeachtete Reize nicht vollständig verloren gehen, sondern „gedämpft“ weiterverarbeitet werden. Im Luftfahrtbeispiel hieße dies, dass Piloten, die Anzeigen im Cockpit überwachen, die Anweisungen des Towers vernehmen

²⁶ [http://www2.hu-berlin.de/psychologie/ingpsycscw/wr2011/pmwiki/pmwiki.php?n= SelektiveAufmerksamkeit.AuditiveAufmerksamkeit](http://www2.hu-berlin.de/psychologie/ingpsycscw/wr2011/pmwiki/pmwiki.php?n=SelektiveAufmerksamkeit.AuditiveAufmerksamkeit) [Zugriff: 28.04.2013].

können und im echoischen Speicher halten, während sich die visuelle Information aus den Cockpitinstrumenten im ikonischen Speicher befindet. Die Aufmerksamkeit kann nach Treisman selektiv auf die verschiedenen Inhalte gerichtet werden, wenngleich einige nur abgeschwächt wahrgenommen werden. Weder von einer frühen Selektion noch von einer Dämpfung gingen hingegen Deutsch und Deutsch (1963) aus. In der **Theorie der späten Selektion** wird angenommen, dass alle Stimuli parallel prozessiert werden, bevor aufgrund semantischer Kriterien entschieden wird, welcher Reiz bedeutsam ist. Im Gegensatz zu dieser Perspektive erfuhr die Dämpfungstheorie von Treisman aber umfangreiche empirische Bestätigung und gilt im Vergleich zu anderen Aufmerksamkeitstheorien als die plausibelste (s. Treisman & Geffen, 1967; Neisser, 1967; Glucksberg & Cowen, 1970).

Ressourcenmodelle postulieren die Verteilung einer Gesamt- oder mehrerer Einzelkapazitäten auf parallele kognitive Prozesse (z. B. Kerr, 1973; Navon & Gopher, 1979; Norman & Bobrow, 1975; Wickens, 1984; Neumann, 1996). Im **Kapazitätenmodell** von Kahneman (1973) wird entgegen den Filtermodellen angenommen, dass Aufmerksamkeitsressourcen auf verschiedene Aufgabenanforderungen nach der Prämisse des kognitiven Aufwandes verteilt werden können: Je höher diese sind, desto größer ist der Ressourceneinsatz. Im Kapazitätenmodell ist Selektion weniger eine Frage des zeitlichen bzw. seriellen Verlaufs der Informationsverarbeitung, sondern eher eine ressourcenbezogene Entscheidung des Operateurs. Kahneman (1973) und später auch Sanders (1983) gingen diesbezüglich von einer übergeordneten **energetischen Komponente (Anstrengung bzw. Effort)** aus, die den Ressourceneinsatz moduliert. Die energetischen Aktivierungsmechanismen bei Sanders bilden die Grundlage für die Informationsverarbeitung: Stimulusprozessierung ist assoziiert mit **Erregung (Arousal)** und motorische Handlungsprozesse sind mit **Aktivierung (Activation)** in Verbindung zu bringen. Entsprechend dürften sich steigende Anforderungen in einer messbaren Erhöhung des Energieeinsatzes niederschlagen (vgl. Roscoe, 1992; Buld, 2000). Auf Basis des Begrenztheitsaxioms können nicht beliebig viele Prozesse simultan verteilt oder beliebig viel Energie eingesetzt werden, sondern es ist mit Leistungseinbußen zu rechnen, wenn die Anforderungen steigen. Wickens (1984, S. 63) fasst zusammen: „*The concept of processing resources is proposed as a hypothetical intervening variable to account for variations in the efficiency with which time-sharing can be carried out; that is, the degree to which two tasks can be performed concurrently as well as each can be performed in isolation*“. Untersucht werden Leistungseinbußen häufig im sogenannten Doppelaufgabenparadigma, in dem Probanden Aufgaben parallel bearbeiten sollen (z. B. verbale Stimuli im Gedächtnis behalten und gleichzeitig visuelle Tracking Aufgabe durchführen; Watkins, Watkins, Craik &

Mazuryk, 1973). Hierin zeigen sich die Kosten durch Mehrfach Tätigkeiten im Vergleich zur seriellen Bearbeitung von Einzelaufgaben.

Basierend auf einer Metaanalyse von mehr als 50 Studien stellte Wickens (2008; vgl. unter anderem 1984) das einflussreiche **4D-Modell multipler Ressourcen** vor (s. Abbildung 16), welches statt nur einer zentralen mehrere Dimensionen der Informationsverarbeitung postuliert. Entsprechend falle die Leistung bei Mehrfachaufgaben dann nicht ab, wenn die Aufgaben verschiedene Ressourcen beanspruchen. Übertragen auf das Beispiel im Cockpit bedeutete dies, dass Piloten problemlos das Control Panel überwachen (visuelle Ressource) und gleichzeitig den Anweisungen des Towers (auditorische Ressource) folgen können. Über die Obergrenze dieser Mehrfacharbeit kann im Rahmen von Wickens Modell aber keine Aussage getroffen werden. Das Modell besteht aus vier Dimensionen:

1. Die Stufen der Informationsverarbeitung (Stages of Processing), worunter die Wahrnehmung, Kognition und die Reaktion auf eine Information zählen. Je nach Stufe können andere Ressourcen beansprucht werden.
2. Die Verarbeitungs-codes (Codes of Processing), zu denen räumliche und verbale Verarbeitung zählen und die jeweils verschiedene Ressourcen beanspruchen.
3. Die sensorischen Modalitäten (Modalities), wozu auditorische und visuelle Wahrnehmung zählen.
4. Die visuellen Kanäle, die in fokales und ambientes Sehen unterteilt werden. Dabei versteht Wickens (2008) unter fokalem Sehen Objektwahrnehmung und komplexe Prozesse wie z. B. Lesen oder Symbolverarbeitung und unter ambientem Sehen die Wahrnehmung und Orientierung im gesamten visuellen Feld (s. dazu Abschnitt 2.2.6).

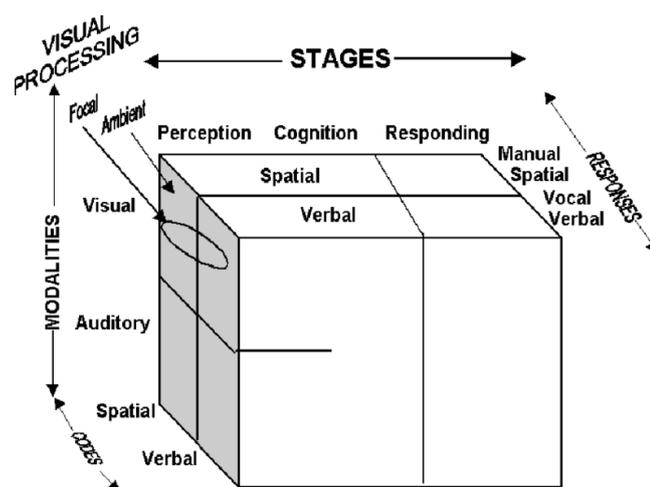


Abbildung 16: 4D-Modell multipler Ressourcen (übernommen aus Wickens, 2008).

Den Zusammenhang zwischen Aufmerksamkeit und Aufgabenanforderungen modellieren Wickens, McCarley und Thomas (s. z. B. 2004) innerhalb des Saliency-Effort-Expectancy-Value-(SEEV)-Modells, welches die Aufmerksamkeitsverteilung in dynamischen Umwelten in einem Bottom Up Prozess, also reizgesteuert, abbildet, der auch Top Down, also kognitionsgesteuerte Prozesse enthält:

$$P(\text{Attend}) = a * \text{Saliency}(S) - b * \text{Effort}(E) + c * \text{Expectancy}(E) + d * \text{Value}(V).$$

Die Bottom Up (Umgebungs-) gesteuerte Aufmerksamkeit wird demnach dann auf einen Reiz gerichtet, wenn dieser salient (Saliency) ist und sie wird durch die Anstrengung (Effort), welche nötig ist, um die Aufmerksamkeit zu verschieben, gehemmt. Beeinflusst wird die Aufmerksamkeit außerdem „Top Down“, durch die Erwartung (Expectancy), einen für das Situationsmodell nützlichen Reiz (Value) zu entdecken (vgl. Wickens et al., 2004). Das SEEV Modell erfuhr umfangreiche empirische Bestätigung in der zivilen Luftfahrt und im Automobilkontext (z. B. Horrey, Wickens, & Consalus, 2006; Wickens et al., 2004). Metz (2009) ergänzt hierzu, dass weitwinklige Blickbewegungen eine höhere Anstrengung erfordern als kleinwinklige und deswegen weniger wahrscheinlich als diese sind. Überträgt man den SEEV-Ansatz auf die Mehrfacharbeit im Cockpit, so ergibt sich: Beim Reiseflug wird die Einhaltung der Flugparameter durch ambiente Wahrnehmung gewährleistet. Für Parallelaufgaben (z. B. Navigation oder Notverfahren) bedarf es fokaler Wahrnehmung. Die Parameter des SEEV-Ansatzes regulieren, wie viel Aufmerksamkeit in Parallelaufgaben investiert wird. Metz (2009, S. 34) betont: *„[Es] kann ein Einfluss der vier genannten Faktoren auf die Ausrichtung der Aufmerksamkeit angenommen werden, die Gewichtung der einzelnen Faktoren ist allerdings unklar.“*

Head Up Displays zur Ressourcenschonung

Der Vorteil kleinwinkliger Blickbewegungen zur strategischen Ressourcenschonung und effizienten Informationsaufnahme wird am Beispiel einer Eye Tracking Untersuchung zu Head Up Displays im Automobilkontext deutlich (Ablassmeier et al., 2007; N = 18). Bei der Bewertung von Head Up und Head Down Displays (HUDs und HDDs) konnte gezeigt werden, dass HUDs eine kürzere Blickdauer erzeugen als HDDs und gleichzeitig die Informationsaufnahme sowie Fahrleistung verbessern. Die Autoren betonen außerdem die Vorteile von HUDs, die sich mit Eye Tracking Daten identifizieren ließen: Sie erzeugen kürzere Reaktionszeiten, gewährleisten eine schnelle Informationsvermittlung, stellen eine höhere Straßenblickzeit sicher und haben eine hohe Nutzerakzeptanz (vgl. auch Weigel et al., 2009). Wie häufig Bottom Up gesteuerte Aufmerksamkeit tatsächlich auf Assistenzsysteme gerichtet wird, ist Operateuren offenbar nicht immer eindeutig bewusst: Um Scanning Techniken und die Gewöhnung an ein Fahrerassistenzsystem (FAS) ab-

zubilden (Sprenger, 2008), wurden objektive Scanning Techniken mit der subjektiven Einschätzung der Probanden verglichen (N = 6). Es zeigte sich, dass die Testfahrer ihre Scanning Techniken falsch einschätzten, das heißt eine Diskrepanz zwischen objektiver und subjektiver Messung vorlag: In Relation zu den tatsächlichen Scanning Techniken unterschätzten die Teilnehmer die Nutzungshäufigkeit des FAS in Kurvenfahrten. Eye Tracking Daten liefern hier also Anhaltspunkte für den sicherheitskritischen Wert von Anzeigesystemen im Cockpit (s. detaillierte Darstellung der Rolle von Sekundäraufgaben in Abschnitt 2.2.8).

Auch in der Luftfahrt ist das HUD ein viel beforschtes Anzeigesystem, z. B. in modernen Kampfflugzeugen. Dixon, Krueger, Rojas und Martin (1990, US Air Force) untersuchten Scanning Techniken für den Luft-Luft-Kampf bei F-15 Piloten (N = 6) im Simulator bezüglich verschiedener AOIs im Cockpit, darunter das HUD. Variiert wurde der Inhalt der Missionen (Offensive, Defensive, Mutual Support, Neutral). Es zeigte sich, dass das HUD am meisten bei offensiven Missionen genutzt wird, bei defensiven weniger. Für defensive Missionen wird eher in periphere Bereiche OTW geblickt. Die durch Eye Tracking entschlüsselten Scanning Techniken zeigen also beispielsweise, dass das HUD ein für offensive Missionen effizientes Informationssystem ist, weil damit die Aufmerksamkeit von Piloten ressourcenschonend kanalisiert werden kann. Dies ist unter anderem darin begründet, dass das HUD einige Head Down Blicke überflüssig macht und damit die visuelle Ressource des Piloten entlastet.

Welchen Beitrag ***bildgebende Verfahren*** und damit neuere Befunde zur Erforschung der Aufmerksamkeitsverteilung leisten, lässt sich am Beispiel der Spotlight Metaphor (z. B. Posner, Petersen, Fox, & Raichle, 1988) verdeutlichen. Die Metapher postuliert, dass Aufmerksamkeit analog zu einem Scheinwerfer gesteuert wird: Die Reaktion auf einen Zielreiz variiert mit dem Ort des Scheinwerferlichts; also dem Fixationspunkt (Posner, 1978; Posner, 1980; Posner et al., 1988; LaBerge, 1983). Mit anderen Worten herrschte bis vor kurzem Konsens darüber, dass Aufmerksamkeit erst auf eine nicht-foveale Region gelenkt werden muss, damit Stimuli hierin identifiziert und verarbeitet werden können. Eine fMRI-Studie, bei der Probanden einen Zielreiz visuell verfolgen mussten, während ihr Hirnstoffwechsel aufgezeichnet wurde, eröffnete allerdings „[...] *that the spotlight metaphor [...] captures only the most salient feature of the attentional topography [...]. The more subtle modulations [...] suggest a gradient of attention [...] but this gradient varies significantly across the field of view. The gradient can extend outward as well as inward from the attended segment depending on its eccentricity*“ (Datta & DeYoe, 2009, S. 7). Für die visuelle Aufmerksamkeitsallokation fassen die Autoren zusammen, dass sie als eine Funktion der Zielreizlokation variiert und verän-

dern die Spotlight Metaphor zum **Window of Attention**. Gemäß den Befunden ist davon auszugehen, dass visuelle Aufmerksamkeit entsprechend der Stimuluslokation aber auch in einem größeren Bereich darum im visuellen Cortex abgebildet wird. Damit ist sie retinotop und kann flexibel auf weitere Reize gerichtet werden. Wie sich diese Flexibilität verhält, wenn weitere, neben der visuellen Ressource angesprochen werden, wird in Zukunft zu klären sein. Die Befunde von Datta und DeYoe unterstützen die Theorie der gedämpften Reizverarbeitung (Treisman, 1960), denn offenbar werden Stimuli auch dann wahrnehmungsbezogen repräsentiert, wenn sie nicht direkt fixiert werden (s. Abschnitt 2.2.6).

2.2.6 Visuelle Aufmerksamkeit

Im Folgenden sollen attentionale Grundlagen von Scanning Techniken im Vordergrund stehen, denn Eye Tracking setzt voraus, „*dass Blickbewegungen nicht der Selektionsprozess selbst sind, sondern das Ergebnis attentionaler Selektionsprozesse*“ (Geise, 2012, S. 7). Auch Henderson (2003, S. 498) sieht Scanning Techniken als messbare Größen der Aufmerksamkeitsallokation: „*[...] vision is an active process in which the viewer seeks out task-relevant visual information [...] and because eye movements are an overt behavioral manifestation of the allocation of attention in a scene, eye movements serve as a window into the operation of the attentional system.*“ Überträgt man diesen Gedanken auf Wickens' (2009) Modell der Informationsverarbeitung von Piloten, ließen sich Scanning Techniken als messbarer Indikator der visuellen Aufmerksamkeit einfügen (s. Abbildung 17). Eine erweiterte Perspektive nehmen Spivey, Richardson und Dale (2008, S. 24) ein, denn sie sehen Scanning Techniken nicht nur als Output- sondern auch als Prozessvariable: „*[...] movements of the eyes to particular locations in space not only provide a measure of cognitive processing, they influence cognitive processing*“ und untermauern die Unterstellung einer Feedbackschleife von Handlungen (hier „aktives“ Scanning) auf die Aufmerksamkeit und Informationsverarbeitung.

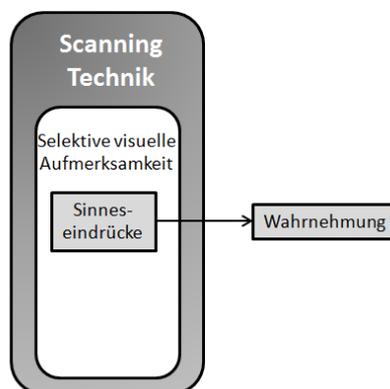


Abbildung 17: Scanning Techniken im Modell von Wickens (2009).

Weniger im Fokus stehen im Folgenden die anatomischen und neurophysiologischen Grundlagen der visuellen Wahrnehmung (für einen Überblick s. Pinel, 2007), sondern vielmehr die theoretischen Hintergründe zur visuellen Verarbeitung und die Rolle von Eye Tracking Ergebnissen bei deren Erforschung. Geise (2012, S. 35) weist darauf hin, dass *„der Eyetracking-Befund [...] lediglich [dokumentiert], ob der Rezipient überhaupt visuelle Aufmerksamkeit auf den Stimulus richtet, wann und wie lange; ob die visuelle Aufmerksamkeit dabei aber endogen kontrolliert ist – und sich der Rezipient bewusst und zielorientiert einem Reiz zuwendet, etwa weil die Aufgabenstellung dies impliziert – oder exogen kontrolliert und damit primär reizgesteuert ist [...], lässt sich aus dem Befund nicht schließen.“* Dieser Hinweis fordert zu einer Auseinandersetzung damit auf, welche Prozesse visuelle Aufmerksamkeit modulieren. Hier treten die bereits erwähnten Vorgänge der exogenen und endogenen Aufmerksamkeitsallokation sowie der fovealen, parafovealen und peripheren Wahrnehmung in den Vordergrund.

Exogene²⁷ Aufmerksamkeitsallokation wird, wie bereits beim SEEV-Ansatz erwähnt, primär aufgrund der Eigenschaften von Stimuli gesteuert; Wahrnehmungsintentionen spielen dabei keine Rolle (Geise, 2012). Die Autorin (ebd.) konnte mit Eye Tracking nachweisen, dass exogene visuelle Aufmerksamkeit interindividuell invariant und sich sogar bei Wiedervorlage des gleichen Stimulusmaterials (unbekannte Wahlplakate mit der Aufforderung, sie einfach zu betrachten) fast identische Blickwege der Probanden zeigen. Ohne konkrete Anforderungen durch die Aufgabenstellung, also z. B. bei reiner Exploration von Bildern, lassen sich durch Eye Tracking kaum interindividuelle Unterschiede zwischen Personen aufdecken.

Im Gegensatz zur exogenen Aufmerksamkeitsallokation, stellt sich die **endogene**²⁸ als intendierter kognitiver Prozess dar, der durch Erwartungen, Absichten und Wissen gesteuert ist (Geise, 2012). Henderson (2003) unterteilt die Quellen endogener visueller Wahrnehmung in a) episodisches Szenenwissen, b) schematisches Szenenwissen und c) aufgabenbezogenes Wissen. Beim episodischen Szenenwissen (a) ordnet der Autor die endogen gesteuerte, visuelle Suche von sowohl kurzzeitig relevanten Umweltreizen (z. B. die Suche des Balls beim Fußballspiel) als auch langfristig gespeicherten Reizlokalitäten (z. B. Ort der Uhr über dem Schreibtisch) ein. Auf schematisches Szenenwissen (b) werde laut Henderson (ebd.) zu-

²⁷ Für die Steuerung der Aufmerksamkeit aufgrund äußerer (exogener) Stimuli existiert eine Vielzahl von Synonymen, wie beispielsweise „Bottom Up Verarbeitung“, „automatische oder automatisierte Verarbeitung“, „präattentive Verarbeitung“ oder „Stimulus-Based Gaze Control“ (vgl. Treisman & Gelade, 1980; Anderson, 1989; Henderson, 2003).

²⁸ Auch für die endogene Aufmerksamkeitsallokation existiert eine Reihe von Synonymen, wie z. B. „Top Down Verarbeitung“, „kontrollierte Verarbeitung“, „attentive Verarbeitung“ oder „Knowledge-Driven Gaze Control“ (vgl. ebd.).

rückgegriffen, wenn man sich an typischen Orten orientiert (z. B. in einer Küche). Bei aufgabenbezogenem Wissen (c) als Quelle endogener visueller Wahrnehmung ordnet der Autor die Durchführung typischer Aufgaben ein (z. B. Blicke in den Rückspiegel bei Spurwechselmanövern).

Blickdauer als Indikator der Spurwechselabsicht

Henning (2010) nutzte die Eye Tracking Methode zur Analyse von Indikatoren des Spurwechsels (Entwicklung eines Spurwechselassistenten; Technische Universität Chemnitz). Dabei konnte festgestellt werden, dass höherfrequente und längere Blicke in den linken Außenspiegel als Indikator dieser Absicht geeignet sind. Der Autor betont aber die Probleme des Indikators: Die Blickdauer in den linken Außenspiegel verlängert sich auch bei „Anwesenheit von Objekten im Spiegel“, ohne, dass ein Spurwechsel ausgeführt wird. Außerdem weist sie eine hohe interindividuelle Varianz auf und könne „teilweise durch Persönlichkeitsmerkmale, wie z. B. die `Inspection-Time`, aufgeklärt werden“ (S. 21). Anhand dieses Beispiels wird einmal mehr deutlich, welche interpretationsbezogenen Herausforderungen Eye Tracking Parameter aufweisen. Scanning Techniken mögen mit Aufmerksamkeitsprozessen assoziierbar sein und in gewisser Weise Informationsverarbeitung abbilden. Inwieweit sie jedoch zur Unterscheidung endogener vs. exogener Prozesse bzw. zur validen Vorhersage von konkreten Intentionen (Spurwechsel) herangezogen werden können, bedarf weiterer Forschung (vgl. ebd.).

Die Kernannahme, die der Eye Tracking Methode zugrunde liegt, ist wie bereits dargestellt (s. Abschnitt 2.1.6), die **Eye-Mind-Hypothese** (Just & Carpenter, 1980; vgl. z. B. Poole & Ball, 2000; Tsai et al., 2007; Metz, 2009; Geise, 2012), wonach der Fixationsort gleichzeitig Ort der Aufmerksamkeitsallokation ist und die Fixationsdauer der Dauer der Informationsverarbeitung eines visuellen Reizes entspricht. Geise (2012, S. 27) zitiert hier eine Reihe von Studien, die implizieren, dass diese Annahme **empirisch belastbar** ist: „Empirische Eyetracking-Befunde zur Analyse kognitiver Verarbeitungsvorgänge [...] bei alltäglichen Aufgaben, etwa beim Autofahren (Land & Lee, 1994), beim Tischtennis- (Land & Furneaux, 1997) oder Cricket-Spielen (Land & McLeod, 2000), beim Zubereiten von Tee (Land, Mennie, & Rusted, 1999) oder Sandwichs (Hayhoe, Shrivastava, Myruczek, & Pelz, 2003), weisen in eine ähnliche Richtung. Hierbei zeigt sich relativ kongruent, dass Fixationen mit dem Verlauf der jeweiligen Aufgabe zeitlich eng verbunden sind (Hayhoe, 2004; Hayhoe & Ballard, 2005), während Objekte, die zur Lösung der jeweiligen Aufgabe irrelevant sind, nicht oder kaum visuell fixiert werden (Hayhoe, Shrivastava, Myruczek, & Pelz, 2003; vgl. Land & McLeod, 2000).“ Zur Validität der Annahme, wonach die Fixationsdauer mit der kognitiven Verarbeitungsdauer assoziierbar ist, existieren ebenfalls Studien, die einen Anstieg der Fixationsdauer in

Abhängigkeit der Komplexität der Aufgabe zeigen (z. B. Velichkovsky, Sprenger & Pomplun, 1997): Sollten rein visuelle Stimuli verarbeitet werden, betrug die Fixationsdauer 120 bis 250 ms, standen jedoch metakognitive Verarbeitungsschritte im Vordergrund, erhöhte sich diese auf bis zu 500 ms.

Die Eye-Mind-Hypothese besitzt aber **nicht in allen Kontexten Gültigkeit**. Anderson, Bothell und Douglass (2004, S. 230) zeigen in einem Gedächtnisabruf-Experiment, dass die Blickdauer nicht in Abhängigkeit der Anzahl von Fakten über Personen oder Orte variiert: „*These results serve to refine the mind-eye hypothesis. Eye movements do not necessarily reflect mental processes, but they do reflect ongoing processes to the extent that the processes depend on the encoding of information. Although some cognitive tasks have to wait for information to be encoded, in many cases, [...], the mind has a path to travel after the encoding of the information, and eye movements do not indicate what is happening on that path.*“ Den Autoren zufolge bedeutet dies also, dass eine längere Blickdauer nicht automatisch auf höhere Informationsverarbeitungsprozesse schließen lässt, sondern zunächst nur eine aufwändigere Enkodierung auf der Ebene der visuellen Aufmerksamkeit angenommen werden kann. Inwiefern z. B. ein länger fixiertes Wort kognitiv tatsächlich verarbeitet (also wahrgenommen) wird, lässt sich aus Eye Tracking Daten nicht ableiten. Das Problem bei Rückschlüssen aus Studien mit Wörtern oder Buchstaben als Stimulusmaterial ist, dass diese nicht unbedingt zum Verständnis der Verbindung zwischen Blickdauer und Kognition im Kontext der Szeneriewahrnehmung oder Hubschrauberführung beitragen. Dennoch zeigen weitere Beispiele, dass sich die Eye-Mind-Hypothese nicht uneingeschränkt anwenden lässt: Hierzu zählt das im Rahmen von Psychotherapie angewendete Verfahren des EMDR (Eye Movement Desensitization and Reprocessing)²⁹, bei dem die Blickrichtung des Patienten nicht mit aktuell wahrgenommenen Umgebungsreizen in Verbindung steht. Das Phänomen der Repetition Blindness³⁰ zeigt außerdem, dass visuelle Reize trotz intendierter Aufmerksamkeit unter bestimmten Bedingungen nicht wahrgenommen werden. Im Repetition Blindness Paradigma kann zwar aus der Blickrichtung auf den Ort der Aufmerksamkeit geschlossen werden (Kanwisher, 1987), nicht jedoch auf gründliche Informationsverarbeitung des visuellen Materials.

²⁹ Hierbei werden Patienten mit Posttraumatischer Belastungsstörung gebeten, im Rahmen einer Therapiesitzung mit ihrem Blick dem sich hin und her bewegenden Finger (ca. 20 bis 40 Mal) des Therapeuten zu folgen. Damit soll eine Stimulation beider Hirnhälften und die Vernetzung von fragmentarischen Gedächtnisinhalten initiiert werden. Eine Metaanalyse zeigt aber, dass die Wirksamkeit von EMDR offensichtlich nicht höher ist als von anderen Expositionstherapiemethoden (Davidson & Parker, 2001).

³⁰ Im Repetition Blindness Paradigma mit schneller serieller, visueller Präsentation von Stimuli (z. B. fünf Wörter pro Sekunde) zeigt sich, dass die Wahrscheinlichkeit dafür, einen wiederholten Zielreiz wahrzunehmen geringer ist gegenüber der Wahrscheinlichkeit, einen neuen Zielreiz wahrzunehmen.

Neben der Tatsache, dass Eye Tracking Rohdaten bisher nicht die Frage nach endogener oder exogener Wahrnehmungssteuerung beantworten können, kann hiermit ebenfalls nicht direkt geklärt werden, welche Rolle Reize außerhalb des fovealen Sehbereiches bei der Aufmerksamkeitssteuerung spielen. Dies wird am Beispiel **parafovealer** (bis zu 5° um die Fovea) bzw. **peripherer Wahrnehmung** (außerhalb des parafovealen Bereiches) innerhalb der Leseforschung deutlich: Gezeigt werden konnte, dass bereits eine „*verdeckte Teilverschiebung visueller Aufmerksamkeit – covert attention – auf das nächste, im parafovealen Bereich liegende Wort vorgenommen werden kann, das dann bereits 'aus dem Augenwinkel' vorverarbeitet wird*“ (Findlay, 2005; Hoffman, 1999; Theeuwes, 1993; zitiert nach Geise, 2012, S. 16). Insofern kann nicht klar zwischen fovealer und parafovealer Wahrnehmung diskriminiert werden, sondern beide sind scheinbar in einer späteren Phase der Verarbeitung integrierbar (ebd.). In ihrer Abhandlung bemerkt Geise, dass peripheres und parafoveales Sehen den Blickweg zwar teilweise steuern, die Prozesse durch Eye Tracking aber nicht abgebildet werden. Entsprechend gilt es, sie im Kontext des Experimentalparadigmas zu interpretieren.

2.2.7 Visuelle Aufmerksamkeit in der Luftfahrt

Das Auge als Sinnesorgan ist für Piloten die **wichtigste Aufmerksamkeitsressource**, denn es liefert Informationen über die Lage im Raum, es ermöglicht das Erkennen von Hindernissen sowie das Abschätzen von Entfernungen und ist damit zentrales Werkzeug für Navigation im Luftraum. Bis zu 80 % aller Informationen werden durch das Auge aufgenommen (EASA, 2010). Neben physiologischen Stressoren (Sauerstoffmangel, Gz-Kräfte) können weitere dysfunktionale Einflüsse „optimale“ Scanning Techniken beeinträchtigen. Hierunter fallen beispielsweise eine zu hohe mentale Belastung oder ein nicht-ergonomisches Design im Cockpit. Diese Belastungsfaktoren stehen nicht nur in Wechselwirkung mit perzeptuellen Prozessen, sondern können Störungen der (visuellen) Aufmerksamkeit, eine Verzögerung kognitiver und psychomotorischer Abläufe sowie Abrufprobleme kurzfristig gespeicherter Gedächtnisinhalte bedingen (Flugmedizinisches Institut der Luftwaffe, 2006, S. 98, 199ff).

Die amerikanische Luftfahrtbehörde Federal Aviation Administration (FAA, 1996) beschreibt unter anderem die kritischen Zustände Absorption, Fixation und Tunneling als Folgen von belastenden Situationsmerkmalen, wobei allen Stadien gemein ist, dass die Aufmerksamkeit des Operateurs gefesselt und dieser mental in einer Situation bzw. seinem Abbild der Situation gefangen ist. Ausgehend davon können diese Zustände Indikatoren für zu hohe Aufgabenanforderungen bzw. eine

inadäquate Ressourcenallokation durch den Operateur sein. Fixation und Tunneling werden dabei synonym verwendet. Unter Visual Tunneling und Tunnel Vision werden Phänomene wie z. B. Einengung des visuellen Aufmerksamkeitsfokus, Einschränkung der Blicke in periphere Bereiche und Abnahme der visuellen Überwachungsfähigkeit subsumiert (vgl. Tsai et al., 2007).

Hinsichtlich dieser kritischen visuellen Phänomene scheinen einsatzbereite Piloten gegenüber der „Normalbevölkerung“ dennoch einen Vorteil aufzuweisen, denn sie sind weniger anfällig als Novizen, da sie aufgrund ihrer Erfahrung und Scanning Techniken ein **größeres funktionelles visuelles Feld** haben (Williams, 1995). Strasburger et al. (2011, S. 57) erklären für das Trainieren der peripheren visuellen Wahrnehmung bei Novizen: *“The reduced perceptual dimensionality of extra-foveal vision is associated with an almost 10-fold increase in learning duration.”* Diesen Nachteil von untrainierten Personen erklären die Autoren mit einer geringeren kortikalen Repräsentation des peripheren visuellen Feldes (ebd., S. 60): *“Cognitive functions in peripheral vision [...] can be characterized in terms of lower representational complexity [...] and processing speed. This might reflect a limited access of the peripheral visual field to prefrontal cortex.”*

Zum „optimalen“ Scanning von Piloten existieren heuristische Empfehlungen der Flugsicherheitsbehörden. Die amerikanische Federal Aviation Administration empfiehlt in der **zivilen Luftfahrt** ein systematisches visuelles Absuchen des Luftraumes (See and Avoid):

Die Empfehlungen zum „optimalen“ Scanning

Laut der FAA (1998)³¹ sollen Piloten

- mindestens 1 Sekunde lang einen Punkt OTW fixieren und dann
- den Blick maximal 10° von diesem Punkt abwenden und erneut mindestens 1 Sekunde OTW fixieren. Außerdem sollen Piloten
- ihren Blick maximal 4 bis 5 Sekunden ins Cockpit richten und
- mindestens alle 16 Sekunden die Umgebung wie oben beschrieben scannen (vgl. Colvin, Dodhia, & Dismukes, 2005).

Eine ähnliche Empfehlung gibt die Europäische Flugsicherheitsbehörde EASA (2010)³² heraus: Piloten sollen

³¹ „The FAA and other organizations recommend a systematic visual search scan for traffic in which the pilot fixates at a location for at least one second, then shifts gaze no more than 10 degrees in order to sequentially scan the entire the visual field outside the window. Pilots are advised to look inside the cockpit no more than 4-5 seconds for every 16 seconds spent scanning the outside world (FAA, 1998a, [...])“ (zitiert nach Colvin, Dodhia & Dismukes, 2005, S. 1).

- die Augen kurz und gleichmäßig bewegen,
- die Bewegungen sollen sich im Bereich um 10° mit jeweils 1 Sekunde Verweildauer befinden,
- der reguläre Blick ins Cockpit sollte maximal 3 Sekunden und
- ein kompletter Außen-Check maximal 20 Sekunden dauern.

Einige Eye Tracking Studien in Flugsimulatoren und unter VFR-Bedingungen aus der zivilen Luftfahrt legen nahe, dass Piloten diesen Empfehlungen nicht folgen, die Blicke OTW zu selten sind (Anders, 2001; Colvin et al., 2005) und sich keine systematischen Scanning Techniken identifizieren lassen (FAA, 1998; Colvin et al., 2005). Williams (2000) beispielsweise zeigt mit Eye Tracking Daten, dass Piloten nur 14 % der Flugzeit OTW geblickt haben (N = 36), wobei die Zeit OTW und die Anzahl entdeckter, anderer Flugzeuge zu $r = .49$ korrelieren. Die Gründe für das Fehlen systematischer Scanning Techniken von Piloten werden in der Literatur nicht explizit diskutiert. Vermutlich erfordert ein konsistentes und wiederholtes OTW Scanning ein gründliches Training (vgl. Colvin et al., 2003) sowie ständige, bewusste Initiierung des Piloten, was die subjektive Belastung erhöht und schneller zu Ermüdung führen kann. Es liegt außerdem die Vermutung nahe, dass Piloten aufgrund von steigenden Anforderungen durch **Glass Cockpits** und damit verbundener **Automation** ihre Head Down Anteile ohnehin erhöhen müssen. Ferner ist denkbar, dass gerade erfahrene Piloten durch ihre Scanning Techniken **Ressourcen schonen** und genau wissen, wann welcher Blick notwendig ist. Colvin et al. (2005) finden einen Anteil von 33 % der Gesamtblickzeit OTW bei Zivilpiloten (N = 12), können jedoch zeigen, dass sich dieser Anteil auf 51 % erhöht, wenn die Beanspruchung steigt (zum Einfluss der Workload auf die Informationsverarbeitung, s. Abschnitt 2.2.8). Dies impliziert, dass gerade erfahrene Piloten ihre Scanning Technik je nach Erfordernissen der Aufgabe adaptieren. Neben der häufig rein deskriptiven Auswertung von prozentualen Blickverteilungen fehlen diesen Arbeiten aber in der Regel eine realitätsnahe Versuchsumgebung (sie verwenden teilweise nur PC-Bildschirme als Simulatoren), inferenzstatistische Vergleiche zwischen Erfahrungsgruppen sowie die tieferegehende Analyse der Blickdauer (vgl. Thomas & Wickens, 2004; Anders, 2001; Kasarskis et al., 2001). Was genau unerfahrene Piloten von Experten lernen können, bleibt daher meist offen.

³² „Effective scanning is accomplished by a series of short, regularly-spaced eye movements that bring successive areas of the sky into the central visual field. Each movement should not exceed 10°, and each area should be observed for at least one second to enable detection. [...] [We] suggest 3 seconds for the instrument check and 20 seconds outside“ (EASA, 2010, S. 12).

Scanning Techniken von Zivilpiloten sind insbesondere in Bezug auf Blicke in die Außenwelt nur schwer mit denen von **Hubschrauberführern** vergleichbar, denn es liegen unterschiedliche Flughöhen und Missionen sowie Unterschiede in der Steuerung der Luftfahrzeuge vor³³. Einen Hinweis darauf, wie sich Scanning Techniken von Hubschrauberpiloten zusammensetzen bzw. wie sie operationalisiert werden könnten, liefert folgender Auszug aus einer Expertenbefragung.

Target Fixations: Intendierte vs. unintendierte Blickfixierungen

Im praktischen Kontext der Heeresfliegerwaffenschule wurde in Gesprächen mit Fluglehrern (N = 6) über Scanning Techniken im Cockpit gesprochen. Einen Teil der Scanning Techniken bilden Blickfixierungen (Target Fixations), also Fixationen auf Objekte oder Instrumente, die in ihrer Dauer länger als ein regulärer Blick sind und die als Indikatoren der taktischen Informationsakquise von erfahrenen Piloten gelten. Die Fluglehrer wiesen darauf hin, dass Target Fixations zwischen erfahrenen und unerfahrenen Piloten differenziert betrachtet werden müssen.

So treten bei Flugschülern eher unintendierte Target Fixations auf, die insbesondere bei hoher Beanspruchung dazu führen, dass wichtige Parameter der Flugsituation übersehen werden. Erklärt wird dies unter anderem mit dem noch nicht ausgeprägten bzw. schulungsbedürftigen peripheren Sichtfeld der Flugschüler, deren parafoveale und periphere Wahrnehmung sich erst durch massierte Übung verbessere. Einsatzbereite Piloten hingegen nehmen durch Target Fixations gezielt Informationen auf und halten das Luftfahrzeug damit auf dem gewünschten Flugpfad, indem z. B. die Größenänderung des retinalen Abbildes von Objektkomplexen der Außenwelt ermittelt und der Hubschrauber über Steuereingaben in einem optimalen Anflugwinkel gehalten werden kann. Target Fixations der Experten sind eine erwünschte Strategie im richtigen Moment, bei der von ausgeprägter parafovealer Wahrnehmung und einem größeren peripheren Sichtfeld profitiert wird. Daher können die Target Fixations von Experten als Maßstab für unerfahrene Piloten herangezogen werden.

Sinngemäß berichten Colvin et al. (2005, S. 5) von empirischen Erkenntnissen bei zivilen Piloten, die ebenfalls intendierte Blickfixierungen einsetzen: *“Scanning the outside world strongly favored looking straight ahead, with many fixations directed only a few degrees to either side. We suspect that many of these fixations*

³³ Nicht unerwähnt sollen an dieser Stelle auch Scanning Techniken von **Kampfpiloten** bleiben, welche man ebenfalls kaum mit denen von Zivilpiloten oder Hubschrauberführern vergleichen kann. Studien hierzu liegen vor allem zur Evaluation von Head Up Displays (HUD) in Kampfjets vor (z. B. Dixon et al., 1990; DiNocera et al., 2007). Diese werden in den Abschnitten der Ressourcenmodelle bei Aufmerksamkeit (Abschnitt 2.2.5) und Aufgabenanforderungen (Abschnitt 2.2.9) vorgestellt. Zusammenfassend zeigen diese Arbeiten, dass bei steigenden Anforderungen (offensive Mission oder steigende Anzahl feindlicher Luftfahrzeuge) die Nutzung des HUD zunimmt.

represent not scanning for traffic but rather the default position for gaze, centered along the central axis of the pilot, the aircraft, and the direction of travel. Gazing mainly straight ahead, coupled with peripheral vision, allows pilots to maintain control of the aircraft."

Target Fixations scheinen also einen Teil der Scanning Techniken erfahrener Hubschrauberpiloten zu bilden und variieren offenbar in verschiedenen Expertisegruppen. Über den Einsatz von Target Fixations bei der Hubschrauberführung liegen bisher keine Untersuchungen vor. Ungeklärt ist daher, in welcher zeitlichen Spannweite sich dabei ein regulärer Blick befindet und ab welchem Schwellenwert von Target Fixations gesprochen werden kann. Der Wertebereich der Blickdauer (Minima vs. Maxima) wird in verschiedenen Forschungsbereichen diskutiert.

Die Diskussion um Untergrenzen... Velichkovsky et al. (2003, S. 314) definieren die zeitlichen Grenzen visueller Verarbeitung mit „*Two Levels of Information Processing*“, worunter sie ein „*Preattentive*“ und ein „*Attentive Level*“ summieren (vgl. Treisman & Gelade, 1980). Bezüglich der Fixationsdauer geben die Autoren an, dass das Preattentive Level mit Fixationen unter 140 ms assoziierbar sei und ein grundlegendes Lokalisieren der Umwelt umschreibe, während das Attentive Level mit längeren Fixationen (> 140 ms) einhergehe und Objektidentifizierung ermögliche. Jacob und Karn (2003, S. 581) definieren eine Fixation, die minimaler Bestandteil eines Blickes ist, wie folgt: „*A relatively stable eye-in-head position within some threshold of dispersion (typically ~ 2°) over some minimum duration (typically 100-200 ms), and with a velocity below some threshold (typically 15-100 degrees per second)*“. Andere Autoren geben andere Untergrenzen an. Tabelle 1 soll die Heterogenität reflektieren, die hinsichtlich der Taxierung einer Untergrenze in verschiedenen Forschungsbereichen (z. B. Lesen vs. Luftfahrt) herrscht.

Tabelle 1: Untergrenzen für die Fixationsdauer.

Quelle	Forschungsbereich	Untergrenze [ms]
Inhoff, Briehl, & Schwartz (1996)	Leseforschung	< 50
Vitu, O'Regan, & Mittau (1990)	Leseforschung	70 – 100
Poole & Ball (2000)	Allgemeine Angabe	100
ISO 15007-1 (2002)	Automobilkontext	120
Velichkovsky et al. (2003)	Allgemeine Angabe	140
Jacob & Karn (2003)	Allgemeine Angabe	100 – 200
Merchant (2001)	Luftfahrtkontext	200

Während in der Leseforschung eine Fixation mit dem Kriterium „eine Silbe/ein Wort wurde erfasst“ gleichgesetzt wird, handelt es sich in anderen Bereichen teilweise um heuristische Angaben der Untergrenzen. Immanenter Bestandteil der meisten Eye Tracking Systeme ist die Voreinstellung einer Untergrenze: Angelehnt an ISO 15007-1 sind z. B. im System Dikablis der Firma Ergoneers GmbH 120 ms eingestellt, wenngleich kein empirischer Konsens darüber festzustellen ist.

...und Obergrenzen. Als Beispiel zur Ableitung eines Maximums für die Blickdauer sei der Ansatz von Wierwille (1993) betrachtet: Der Autor ging innerhalb des Paradigmas der visuellen Nebenaufgabe beim Fahren davon aus, dass die Blickdauer auf ein Navigationssystem durch zwei Faktoren determiniert ist. Zum einen sei sie beeinflusst durch die Schnelligkeit der Informationsaufnahme durch den Fahrer, zum anderen determiniere sein Bedürfnis, die Kontrolle über die Fahrsituation herzustellen, wie lange er darauf blickt. Wierwille (ebd.) nahm an, dass nach spätestens 1600 ms dieses Kontrollbedürfnis so virulent sei, dass der Blick vom Navigationssystem auf die Straße zurückkehre. Eine andere Klassifikation der Blickdauer schlagen Zwahlen, Adams und DeBald (1988) vor. Reguläre Blickabwendungen können hiernach maximal 1200 ms dauern, wobei zur Aufgabenerfüllung maximal drei Blicke erforderlich sein sollen. Die Autoren definieren weiterhin eine Grauzone bis zu einer kritischen Blickdauer von 2000 ms. Bedarf eine Nebenaufgabe mehr als vier Blicke, die jeweils länger als 2000 ms dauern (s. Abbildung 18), handelt es sich um eine sicherheitskritische Aufgabe (z. B. Adresseingabe in ein Navigationssystem während des Fahrens).

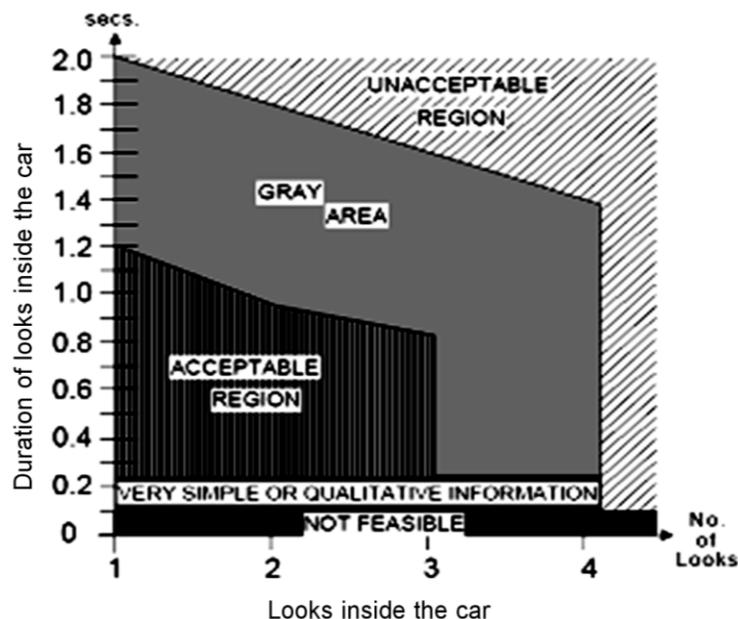


Abbildung 18: Klassifikation der Blickdauer nach Zwahlen et al. (1988).

Dass auch die Diskussion über ein Maximum der Blickdauer im Fahrkontext bisher ohne Konsens blieb, illustriert Tabelle 2.

Tabelle 2: Obergrenzen für die Fixationsdauer (nur im Fahrkontext).

Quelle	Obergrenze [ms]
Theofanou (2002)	1500
ISO 16673 (2006)	1500
Wierwille (1993)	1600
Zwahlen et al. (1988)	2000
Schweigert (2002)	2000
Alliance of Automobile Manufacturers (2002)	2000

Inhoff und Radach (1998, S. 37) weisen darauf hin, dass einige Forscher keine absoluten Cutoff-Werte ansetzen. Es sei zum Teil üblich, solche Daten als Extremwerte zu betrachten, die relativ zur Datenlage zwei bis drei Standardabweichungen nach unten oder oben vom Mittelwert der Blickdauer abweichen. Da sich bisher aufgrund fehlender Standards kaum klare zeitliche Grenzen für Target Fixations ableiten lassen, könnte letztgenannter Gedanke zu einer Grenzwertbestimmung herangezogen werden: Wenn die Blickdauer von Piloten mit einer Standardabweichung SD um einen Mittelwert M streut, dann liegen die Blicke X mit

- $(M_x - 3*SD) \leq X \leq (M_x + 3*SD)$ im Normalbereich dieser Piloten,

während alle Blicke

- $X > (M_x + 3* SD)$ Extremwerte darstellen,

die um ein hohes Maß nach oben vom Mittelwert abweichen und Target Fixations implizieren (abgeleitet aus ebd.). Dieser Ansatz ist zumindest für die ad hoc Bestimmung eines in Relation zur jeweiligen Stichprobe stehenden Cutoff-Wertes gewinnbringend. Ein vergleichbares Vorgehen beschreibt Geise (2012) und schlägt vor, dass die Blickdauer in Klassen eingeteilt werden kann, um danach den prozentualen Anteil der Klassen an allen aufgezeichneten Blicken zu errechnen.

Da die Blickdauer zentral für die Analyse der Scanning Techniken von Hubschrauberpiloten ist und im Kontext der vorliegenden Arbeit unter verschiedenen Aufgabenanforderungen und Expertisestufen betrachtet werden soll, werden im Folgenden theoretische Grundlagen dieser Konzepte einer näheren Betrachtung unterzogen. Hierzu wird zunächst die Rolle des Arbeitsgedächtnisses bei der Bewältigung von Aufgaben und zugehöriger Informationsverarbeitung näher beleuchtet.

2.2.8 Arbeitsgedächtnis, Belastung, Beanspruchung und Workload

Kehrt man nach Betrachtung der Aufmerksamkeit und visuellen Wahrnehmung zum Grundmodell der Informationsverarbeitung (Wickens, 2009) zurück, gelangen sensorische Informationen, beispielsweise Daten aus Cockpitinstrumenten oder Funksprüche, zur Weiterverarbeitung ins **Arbeitsgedächtnis** des Piloten (s. Abbildung 19). Die Konzepte des **ikonischen und echoischen Speichers**, wo visuelle und auditorische Information angesammelt wird, sowie der **zentralen Exekutive**, die die Aufmerksamkeitsallokation und Stimulusselektion aus diesen Speichern steuert, bilden eine Klassifikation des Arbeitsgedächtnisses (Baddeley & Hitch, 1974). *“The term working memory refers to a brain system that provides temporary storage and manipulation of the information necessary for such complex cognitive tasks as language, comprehension, learning, and reasoning”* (ebd., S. 311).

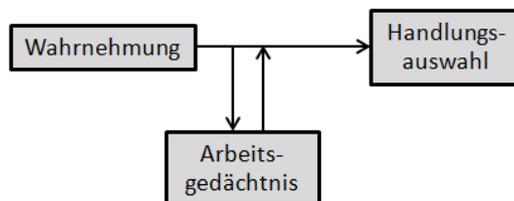


Abbildung 19: Navigation im Modell von Wickens (2009): Arbeitsgedächtnis.

Die Tatsache der Ressourcenbegrenztheit bei der Informationsverarbeitung schlägt sich in Konkordanz zu den Aufmerksamkeitstheorien in theoretischen Ansätzen zum Arbeitsgedächtnis nieder: *“Working memory [...] may be especially sensitive to manipulations in workload because it is thought to be of limited capacity”* (Goldberg et al., 1998, S. 296). Empirische Evidenz liefern neuere Studien mit **bildgebenden Verfahren**. Die Gruppe um Goldberg (1998) untersuchte die Leistung des Arbeitsgedächtnisses mithilfe einer Primär- und Sekundäraufgabe (Kartensortiertest mit auditorischer Beschattungsaufgabe). Unter der Bedingung der parallelen Aufgabebearbeitung zeigte sich, dass der Anstieg der mentalen Beanspruchung zu abnehmender präfrontaler Aktivität – also dem hirnanatomischen Korrelat der zentralen Exekutive – führt, während die temporale Aktivität zunahm. Die Abnahme der präfrontalen Aktivität interpretieren die Autoren als Kausalfaktor für Leistungseinbußen aufgrund inadäquater Handlungsauswahl bei steigender **Beanspruchung** durch die Aufgabe. Was aber verbirgt sich hinter den Konzepten Beanspruchung durch eine Aufgabe, Workload sowie Leistung und wie stehen sie mit Scanning Techniken von Piloten in Verbindung?

Mit dem Begriff Beanspruchung ist der Begriff **Belastung** assoziiert, wobei beide Konstrukte seit den 40er Jahren im deutschsprachigen Raum im Bereich der Arbeitspsychologie, die Arbeitsbedingungen als äußere, auf den Menschen ein-

wirkende Faktoren versteht (Schuler, 2007, S. 161), erwähnt und erforscht werden (für einen Überblick s. Schönplflug, 1987 oder Buld, 2000). Dabei ist die semantische Trennung der Begriffe schwierig: „Mit zunehmender Forschung auf diesem Gebiet scheint es jedoch, als ob die Begriffe eher weniger fassbar als konkreter werden“ (Buld, 2000, S. 27). Uneinigkeit herrscht vor allem hinsichtlich der Gruppierung der Konzepte entweder als Wirkmechanismen im Reiz-Reaktionsmodell oder aber als Konsequenzen von Arbeitstätigkeiten: „Die Zweideutigkeit der Umgangssprache, mit ‚Beanspruchung‘ und ‚Belastung‘ sowohl einen Vorgang als auch einen Zustand zu bezeichnen [...], setzt sich im wissenschaftlichen Sprachgebrauch fort“ (Schönplflug, 1987, S. 133).

Rohmert und Rutenfranz (1975, S. 8, zitiert nach Buld, 2000, S. 28) versuchen, die Begriffe definitorisch aus objektiver vs. subjektiver Perspektive des Individuums zu diskriminieren: „Belastungen [...] sind alle objektiven, von außen auf den Menschen einwirkenden Größen und Faktoren. Beanspruchung sind die subjektiven Folgen der Belastung, deren Auswirkungen im Menschen auf den Menschen“. Subjektive Beanspruchung kann sowohl psychisch – hier sowohl mental als auch emotional – als auch physisch sein. Bei der Aufgabenausführung kann es zu einer Rückkoppelung der Beanspruchung auf die Merkmale des Operators kommen (s. Abbildung 20). Dauerhafte Unterforderung beispielsweise kann zu Fähigkeitsverlust oder emotionale Überbeanspruchung zum Absinken des Leistungsmotivs führen (Rohmert, 1984).³⁴

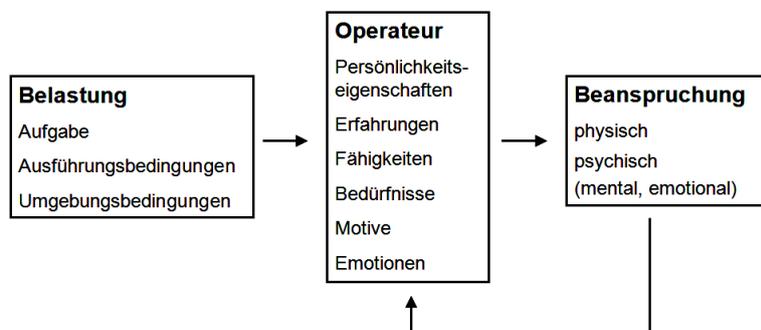


Abbildung 20: Zusammenhang Belastung und Beanspruchung (modifiziert aus Grandt, 2004).

Die relativ griffige Unterteilung der Begriffe nach objektiven und subjektiven Faktoren schlägt sich später in der ISO Normierung nieder (ISO 10075-1 und DIN

³⁴ Herczeg (2005) nennt Folgen der Nichtbewältigung von Beanspruchungen „Beeinträchtigungen“ und zählt hierzu z. B. Frustration, Ärger, nachlassende Leistungen, Angst, aber auch psychosomatische und chronische Erkrankungen. Weitere Beeinträchtigungen nennt Schönplflug (1987), wie z. B. Stress, Fehlbeanspruchung, Sättigung oder Belästigung. Wie Buld (2000, S. 28) klarmacht, bereitet die Einordnung bzw. Abgrenzung dieser Begriffe in das Belastungs-Beanspruchungskonzept Mühe, weil insbesondere Stress nicht eindeutig als objektive Aufgabenanforderung bzw. subjektive Auswirkung auf den Menschen klassifizierbar ist.

33405; zitiert aus Herczeg, 2005, S. 24): „Psychische Belastung wird verstanden als die Gesamtheit der erfassbaren Einflüsse, die von außen auf den Menschen zukommen und auf ihn psychisch einwirken. [...] Belastungen, die für das jeweilige Individuum spürbar werden, bezeichnen wir auch als Beanspruchungen. Psychische Beanspruchung wird verstanden als die individuelle, zeitlich unmittelbare (nicht langfristige) Auswirkung der psychischen Belastung im Menschen in Abhängigkeit von seinen individuellen Voraussetzungen und seinem Zustand.“

Belastungen von Piloten der Bundeswehr (zitiert nach Buld, 2000)

Die Militärpiloten, die in der Studie von Buld hinsichtlich ihres Belastungsempfindens befragt wurden, zitiert die Autorin (S. 83f) wie folgt: „Wenn man beim Autofahren nicht mehr kann, fährt man an den Straßenrand und macht eine kurze Pause. Beim Fliegen kann man nicht anhalten, man muss durchhalten, auch wenn man eigentlich nicht mehr kann.“ Neben dem „Durchhalten müssen“ schlägt das Ertragen von Monotonie zu Buche: „Erst passiert lange Zeit gar nichts und man muss sich auf die reine Überwachungstätigkeit konzentrieren. Und wenn man dann so eingulmt ist, muss man von 0 auf 100, um z. B. Fallschirmspringer abzusetzen.“ Im letzten Beispiel wird deutlich, dass insbesondere die adäquate bzw. ausgewogene Verteilung von Belastungen über die Arbeitszeit hinweg entscheidend für das Beanspruchungsempfinden ist (vgl. Rubio et al., 2004), im fliegerischen Bereich aber häufig das Gegenteil auftritt.

Neben dem deutschen Belastungs-Beanspruchungskonzept ist das englische Konzept **Workload** seit den 60er Jahren mit einem Forschungsstrang verbunden, der eher pragmatisch ist, sich explizit auf die Luftfahrt bezieht und auf die Trennung von Belastung und Beanspruchung verzichtet (Buld, 2000; Flemisch, 2001). Auf den Bereich der Anforderungen im Cockpit lässt sich das Workload-Konzept im Vergleich besser übertragen, weil die Flugführungsaufgabe als Mensch-Maschine-Interaktion eben gerade aus Mehrfach Tätigkeiten mit hohen kognitiven Anforderungen besteht, denen ein Informationsverarbeitungsprozess zugrunde liegt, der durch „[t]he perceived relationship between the amount of mental processing capability or resources and the amount required by the task“ (Hart & Staveland, 1988, S. 2; vgl. Buld, 2000, S. 30; Wickens & Hollands, 1999; Flemisch, 2001; Rubio et al., 2004; vgl. auch Gopher & Donchin 1986, S. 41-43) determiniert wird. Buld (2000, S. 32) weist darauf hin, dass die Datenebene einer Untersuchung darüber entscheidet, in welcher der beiden vorgestellten Konzeptualisierungen die Beanspruchungsmessung durchgeführt wird. Die Autorin assoziiert mit der Workload-Messung eher ein operatorsbezogenes Vorgehen, wobei die Belastungsmessung eher aufgabenbezogen ist. Da in dieser Arbeit die visuellen Ressourcen des Piloten

und nicht die Aufgabeninhalte als abhängige Variablen im Vordergrund stehen, wird hier auf die subjektive Workload der Operateure referiert.

Ähnlich wie beim Belastungs-Beanspruchungskonzept herrscht in Bezug auf Workload Uneinigkeit darüber, ob das Konzept eher Wirkmechanismus oder Folge von variierenden Aufgabenanforderungen ist. So wird Workload in folgender Definition mit anderen Konstrukten gleichgesetzt, die genauso gut Konsequenzen zu hoher Workload sein können: „*At a simplified level, workload can be defined as the cost of accomplishing a task for a human operator. These costs can be fatigue, stress, and errors among others*“ (Hart, 2006, S. 1). Andere Quellen definieren insbesondere die mentale Workload analog zur klar begreifbaren physischen Anstrengung: „*[...] mental workload is the mental analogy to the easily understood concept of physical workload*“ (Vidulich, 2002, S. 116) bzw. „*[...] the effort invested by the human operator into task performance*“ (Hart & Wickens, 1990, S. 258). Daneben existieren Umschreibungen, die darauf hinweisen, dass das Konstrukt vielen Einflussgrößen unterliegt und nicht nur mit physischen Anforderungen verglichen werden kann (z. B. „*[...] mental workload is a multidimensional concept involving an interaction between pilot, task, and environment [...]*“, Svensson et al., 1999, S. 3; vgl. auch DeWaard, 1996). Außerdem gibt es Perspektiven bezogen auf Anforderungen als Determinanten für mentale Workload, wobei Workload hier als reiner Leistungsparameter gesehen wird, das heißt als Konsequenz der Anforderungen, nicht jedoch als Wirkmechanismus: „*Cognitive workload may be characterized as the difference between predicted performance, as based on an easy version of a task, and actual performance, as based on a manipulation that makes the easy version of the task more difficult [...]*“ (Goldberg et al., 1998, S. 296). Zur Messung der Workload lassen sich zwei größere Kategorien unterscheiden (O’Donnell & Eggemeier, 1986; DeWaard, 1996): Auf der einen Seite stehen subjektive (Selbstreport), auf der anderen objektive Methoden, die in Leistungsmessung (Performance) und physiologische Verfahren unterteilbar sind.

Bei den **subjektiven Verfahren** soll die Workload typischerweise durch den Operateur auf einer oder verschiedenen Dimensionen quantifiziert und häufig gewichtet werden. Beispiele sind (vgl. z. B. DeWaard, 1996; Buld, 2000; Rubio et al., 2004; Cain, 2007 spezifiziert 19 subjektive Verfahren zur Workload-Messung):

- Subjective Workload Assessment Technique (SWAT; Reid & Nygren 1988),
- Rating Scale Mental Effort (RSME; Zijlstra & van Doorn, 1985),
- NASA Task Load Index (NASA-TLX; Hart & Staveland, 1988) und
- Workload Profile (WP; Tsang & Velazquez, 1996).

Welche Bewertungsskala angewendet wird, hängt davon ab, welche Information erfasst werden soll: Bei globalen Aussagen über eine Overall-Workload eignen sich unidimensionale Skalen (z. B. RSME), während mehrdimensionale Instrumente ein detaillierteres Bild über die Ausprägung des Ressourceneinsatzes liefern (DeWaard, 1996). Der Selbstreport „*is the most used and probably the most reliable method for assessing cockpit workload*“ (Roscoe, 1992, S. 259), denn *“No one is able to provide a more accurate judgment with respect to experienced mental load than the person concerned”* (DeWaard, 1996, S. 31). Weitere Vorteile der Verfahren liegen in ihrer geringen Intrusivität und ihrer Anwendungsökonomie. Nachteile werden in der Intro- bzw. Retrospektion durch die Operateure und die damit verbundenen Antwortverzerrungen gesehen (ebd.). Außerdem liefern die subjektiven Methoden zwar Auskünfte über die Ursprünge mentaler Workload, erlauben jedoch keine Aussagen in Bezug auf spezifische Aufmerksamkeitsressourcen für eine einzelne Aufgabe (Rubio et al., 2004).

Im Vergleich der Verfahren untereinander ist der **NASA-TLX** sowohl in der zivilen (z. B. Rubio et al., 2004) als auch militärischen (s. RTO/NATO, 2001) Forschung am weitesten verbreitet, da er die höchste prädiktive Validität bei der Vorhersage der Performance zeigt, sensitiv ist für Veränderungen der mentalen Workload und bei der Bewertung von verschiedenen Aufgaben eingesetzt werden kann (DeWaard, 1996; Buld, 2000; Rubio et al., 2004 liefern einen Überblick über die Einsatzbereiche). Der RTO/NATO Technical Report 21 (2001) als Darstellung aus dem militärischen Bereich fasst die Vor- und Nachteile des NASA-TLX zusammen und ist mit zivilen Befunden vergleichbar. Neben bereits dargestellten **messmethodischen Vorzügen** werden die hohe Augenscheinvalidität, die Intervallskalierung, die Vermeidung von Bodeneffekten und die verschiedensprachigen Versionen des Verfahrens genannt. Gleichzeitig werden übersetzungsbezogene Differenzen als **Nachteile** aufgeführt, genau wie die Instruktionen, die unverständlich für Probanden sein können, die Interkorrelation der Subskalen, die Retrospektivität des Verfahrens sowie die mangelnde Relevanz einiger Subskalen für Workload (z. B. Frustration oder physische Anforderungen). Insgesamt erfasst das Instrument die subjektive Workload auf sechs Dimensionen: Mentale, physische und zeitliche Anforderungen sowie Leistung, Anstrengung und Frustration. Als Antwortformat stehen 20-stufige bipolare Skalen von 0 bis 100 % zur Verfügung. Zur Kalkulation eines Gesamtwertes kann optional ein Gewichtungsalgorithmus durchgeführt werden, bei dem der Operateur einen Paarvergleich der Workload-Komponenten vornehmen soll (s. hierzu Hart & Staveland, 1988). Wie bereits Buld (2000, S. 63) anmerkt, existieren Hinweise dafür, dass die deutsche Übersetzung des NASA-TLX kritische Auswirkungen auf die Gütekriterien hat. Daneben zitiert Buld eine Studie von Moroney, Biers und Eggemeier (1995) und konstatiert:

„[Man] kann davon ausgehen, dass das Event-Rating [innerhalb des NASA-TLX] auch ohne Gewichtungungsverfahren zu einer ähnlichen Vorhersage des Gesamtwertes der Workload führt. [Die Autoren] berichten von verschiedenen Studien, in denen die Ergebnisse des NASA-TLX mit und ohne Gewichtung verglichen wurden. Die Zusammenhänge zwischen .94 und .977 sprechen dafür, dass beide Verfahren zum selben Workload-Score gelangen.“ Somit kann der Gesamtwert auch über die Addition aller Skalenwerte und anschließender Division durch sechs errechnet werden (vgl. Byers, Bittner, & Hill, 1989; Fairclough, 1991; DeWaard, 1996).

Im Gegensatz zur relativ zeitökonomischen Anwendung des NASA-TLX präsentieren sich andere Verfahren mit mehreren Durchführungsschritten, die die Intrusivität der Methoden steigern und ihre Augenscheinvalidität mindern: Bei der **SWAT** müssen z. B. in der Phase Development erst alle 27 möglichen Kombinationen von Workload-Faktoren (Time Load, Mental Effort Load, Psychological Stress Load mit jeweils drei Ausprägungen) kompiliert werden, bevor die eigentliche Bewertung in wiederum zwei Schritten durchführbar ist. Rubio et al. (2004, S. 69) halten fest: „*The scale development phase of SWAT was quite tiring for the subjects*“. Zeitaufwändig, wenngleich theoretisch sehr fundiert, zeigt sich auch das **WP**, welches die Dimensionen aus dem Ressourcenmodell von Wickens (1984) abfragt (vier mal zwei Dimensionen, s. Tsang & Velazquez, 1996). Probanden sind hier angehalten, die Aufgaben auf den einzelnen Dimensionen mit einer Zahl zwischen 0 = „die Aufgabe stellte keine Anforderungen“ und 1 = „die Aufgabe erforderte maximale Aufmerksamkeit“ zu bewerten. Aus der Summe der Bewertungen wird dann ein Gesamtwert errechnet. Problematisch dürfte sich hier vor allem der Schritt erweisen, Probanden die Dimensionen aus dem Ressourcenmodell zu erklären, sodass diese die Subskalen distinkt bewerten können. Wie die Autoren einräumen, muss das Verfahren seine empirische Belastbarkeit und diagnostischen Eigenschaften erst noch unter Beweis stellen (vgl. auch Rubio et al., 2004). Neuere Studien, die das WP einsetzen, existieren gemäß Recherchestatus dieser Arbeit noch nicht. Nach Abwägung der dargestellten messmethodischen und ökonomischen Eigenschaften der subjektiven Workload-Verfahren wurde für diese Studie der NASA-TLX favorisiert.

Neben subjektiven Verfahren kommen auch objektive Techniken zum Einsatz, wie beispielsweise die **Leistungsmessung bei der Hauptaufgabe (Primary Task)**, bei der es auf die Akkuratheit der Aufgabenerfüllung ankommt. Typische Leistungsparameter sind Reaktionszeit, Präzision oder Fehlerfrequenz, wobei grundsätzlich die Abweichung von einem Sollzustand (z. B. Halten einer bestimmten Flughöhe) als Indikator dient (Buld, 2000). In der Ausbildungspraxis von Piloten erfolgt die Leistungsbewertung aber nicht immer tatsächlich objektiv (s. Exkurs).

Leistungsbewertung bei der Hubschrauberführergrundausbildung

Im Rahmen eines halbstandardisierten Experteninterviews (N = 6) zählten Experten der Hubschrauberführung aus der Heeresfliegerwaffenschule bei der Frage, welche Leistungsparameter eines Flugschülers nach einer Mission bewertet werden, auf: Die mentale Restkapazität des Schülers nach der Mission (in diesem Sinne Anteil der aufgewendeten mentalen Ressourcen), die geschätzte Abweichung vom optimalen Landepunkt (in Meter) und der Gesamteindruck vom Schüler, auch genannt „Airmanship“. Hierunter wird die sichere und vorausschauende Luftfahrzeugführung und Kommunikation mit anderen Crew-Mitgliedern verstanden. Auffällig ist hierbei, dass die Leistungsbewertung subjektiv durch Fluglehrer erfolgt und in der Heeresfliegerwaffenschule keine objektiven Parameter der Hauptaufgabe (z. B. Reaktionszeit, Präzision in Form von PC-dokumentierten Abweichungsquadraten von der optimalen Flugroute, Fehlerfrequenz) oder die objektive Bewertung einer Zweitaufgabe verwendet werden. Hier besteht auf der einen Seite Standardisierungsbedarf, da die Leistungsbewertung der Flugschüler durch subjektive Einflüsse der Fluglehrer verzerrt sein kann. Der Erforschung der Interraterreliabilität bei der Leistungsbewertung von Flugschülern sollte daher in Zukunft erhöhte Aufmerksamkeit geschenkt werden.

Auf der anderen Seite zeigen Studien, dass die Akkuratheit von Leistungsbeurteilungen durch Experten (z. B. Lehrer) in Bezug auf die Leistung (von z. B. Schülern) häufig sehr hoch ist. Coladarci (1986) weist nach, dass Schullehrer ihre Schüler in verschiedenen Kompetenzbereichen mit einer Akkuratheit von $.67 \leq r \leq .85$ und damit relativ adäquat bezüglich der objektiven Leistung bewerten. Jako und Murphy (1990) zeigen, dass eine Dekomposition (das heißt die Unterteilung der Bewertung in mehrere Bereiche) zu einer höheren Akkuratheit der subjektiven Leistungsbewertung führt. Diesem Prinzip wird in der HGA durch die Bewertung von den drei genannten Bereichen (Restkapazität, Abweichung, Airmanship) Rechnung getragen. Außerdem wird durch Studien zur Interraterreliabilität klar, dass Expertenbewertungen untereinander eine hohe Konsistenz aufweisen ($r = .97$ bei Borman, 1977; $r = .80$ bei Akinwuntan, DeWeerd, Feys, Baten, Arno, & Kiekens, 2003, zur Bewertung der Fahrleistung).

Betrachtet man die Akkuratheit bei Selbstbewertungen von Experten, spielen weitere Einflüsse eine Rolle. Etablierten psychologischen Befunden zufolge sind Selbstbewertungen der eigenen Leistung häufig selbstwertdienlich verzerrt (vgl. Stroebe, Jonas, & Hewstone, 2002). Es werden beispielsweise keine Korrelationen zwischen den Selbstbewertungen von Ärzten hinsichtlich ihres Fachwissens und tatsächlich gezeigten Kenntnissen in Leistungstests gefunden. Analog schätzen Krankenschwestern ihr Wissen über lebensrettende Maßnahmen höher ein

als es in Anwendungssituationen tatsächlich ist. In einer anderen Studie zeigt sich, dass Waffenbesitzer sich selbst hinsichtlich Waffennutzung, Sicherheit und Fachkenntnis überschätzen (Beispiele nach Ehrlinger & Dunning, 2003). Insgesamt schwankt die empirische Akkuratheit der selbstbewerteten Leistung sehr stark und in Abhängigkeit des Experimentaldesigns. Moorthy, Munz, Adams, Pandey und Darzi (2006) finden eine akzeptable Akkuratheit von $r = .64$ bei der Selbstbewertung der Leistung von medizinischen Spezialisten bei einer simulierten Operation, wenn die Items zur Selbstbewertung detailliert auf die Kriterien der Aufgabe zugeschnitten sind. Auch bereits Mabe und West (1982) identifizieren in ihrer Metaanalyse von 55 Studien in vielen Bereichen (schulische, berufliche und sportliche Leistung) signifikant positive Korrelationen zwischen Selbsteinschätzung und objektiver Leistung (mittlere Akkuratheit $r = .29$). Die Autoren extrahieren Faktoren, die die Präzision der Selbstbewertung steigern können:

- die Selbstbewertung wird mit objektiven Daten in Verbindung gebracht,
- Probanden haben Erfahrung mit Selbstbewertungen,
- bei der Selbstbewertung wird Anonymität garantiert und
- die Instruktion zur Selbstbewertung betont den Vergleich mit Anderen.

In einer aktuelleren Studie zur Akkuratheit von Selbstbewertungen zeigen Ehrlinger und Dunning (2003), dass außerdem der Faktor der Metakognition bzw. das „permanente Selbstbild“ (Chronic Self-View) über die eigene Leistung zur Präzision der subjektiven Leistungsbewertung beitragen kann: Die Studie beschäftigt sich mit der selbsteingeschätzten Leistung bezüglich des wissenschaftlichen Denkvermögens (Scientific Reasoning) und deckt eine Akkuratheit von $r = .61$ bis $.68$ auf, wenn der moderierende Einfluss der Metakognition hierüber berücksichtigt wird. Schlussfolgernd dürfte für Hubschrauberführer zumindest theoretisch gelten, dass Selbstbewertungen der fliegerischen Leistung dann akkurater sind, wenn diese über ein höheres Maß an Metakognition über die eigene Leistung verfügen. Dass diese Annahme im Hinblick auf erfahrene Hubschrauberpiloten gerechtfertigt ist, wird in Abschnitt 2.2.10 näher beleuchtet.

Neben der Leistungsbeurteilung in der Hauptaufgabe ist es möglich, die **Leistung bei einer Zweitaufgabe (Secondary Task)**, die parallel bearbeitet wird (z. B. Reaktion auf Warnanzeigen), zu erfassen. Dabei wird davon ausgegangen, dass erhöhte Workload zu abnehmender Präzision der Aufgabenerfüllung führt (DeWaard, 1996; Buld, 2000). Dem Zweitaufgabenparadigma liegt die Annahme zugrunde, dass Ressourcen, die in der Erstaufgabe nicht gebraucht werden, für die Bearbeitung der Zweitaufgabe herangezogen werden können. Steigen die Anforderungen der Erstaufgabe, ist mit einem messbaren Leistungsabfall in der Zweitaufgabe zu

rechnen (Vidulich, 2002). Die Aufgabenanforderungen können dabei zweierlei Ausprägungen erreichen: Im Optimalfall sind sie geringer als die zur Verfügung stehenden Ressourcen des Operateurs, sodass dieser im Falle unerwarteter Ereignisse freie Kapazitäten hat. Im Extremfall kommt es aber zum Zusammenbruch der Mensch-Maschine-Interaktion und zum Leistungsabfall, weil die Anforderungen der Aufgabe die Ressourcen des Operateurs übersteigen (Navon & Gopher, 1979). Einige Autoren sprechen hier von der „Red Line der Workload“ (z. B. De-Waard, 1996; Grier, 2008). Flemisch (2001) fasst den Zusammenhang zwischen Workload und Leistung (aus Wickens & Hollands, 1999) wie folgt zusammen (s. Abbildung 21): Region A repräsentiert den Optimalfall, in dem mehr Ressourcen verfügbar als gefordert sind. In Region B und C ist die Workload höher als die Ressourcen des Operateurs; es kommt zu einem Abfall der Leistung. Ausgegangen wird von einer approximativ gegen ein hypothetisches Maximum strebenden Reservekapazität und Leistungsfähigkeit des Operateurs.

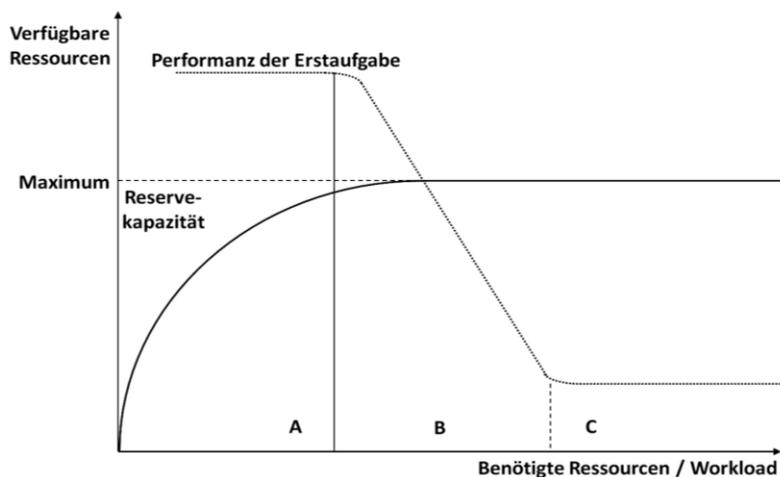


Abbildung 21: Verfügbare vs. benötigte Ressourcen (übernommen aus Flemisch, 2001).

Für den Fall, dass mehr Ressourcen verfügbar als gefordert sind (Region A, Abbildung 21), muss aber nicht uneingeschränkt gelten, dass sich die Leistung eines Operateurs approximativ einem Maximum nähert. Wie bereits im Yerkes-Dodson-Gesetz (1908) beschrieben, kann die Leistung bei zu geringem Aktivierungsgrad abfallen; der Leistungs-Erregungs-Zusammenhang ist mithin nicht linear. Johannsen (1979) überträgt den umgekehrt-U-förmigen Zusammenhang auf das Verhältnis von Beanspruchung und Leistung: Ein suboptimales Maß an Beanspruchung (Unter- oder Überforderung) kann zu Fehlern und einer unakzeptablen Performance führen. Auch DeWaard (1996, S. 21) erweitert den Zusammenhang – „it is not clear where the domain of underload is“ – und ergänzt eine Region D, wo er insbesondere monotone Aufgaben einordnet: „These are low demand tasks that can result in increases in task difficulty and workload by a reduction in capacity“.

Workload und Visual Tunneling im Automobilkontext

Mit dem Sekundäraufgabenparadigma (Fahren mit auditiver Nebenaufgabe) analysieren Tsai et al. (2007) den Visual Tunneling Effekt (N = 13) mit Eye Tracking. Die Autoren variierten die Anforderungen der Aufgabe (Verkehrsdichte) und zeigten unter anderem, dass die Spurabweichung (nachlassende Fahrleistung), das Visual Tunneling (prozentualer Anteil von Blicken), die horizontale Vergenz (unwillkürliche Blickbewegungen) und die Lidschlagfrequenz bei steigender Beanspruchung zunehmen. Einen Anstieg der Fixationsdauer auf ein FAS bei steigender Gefahrenwahrnehmung, eine Verkleinerung des horizontalen Blickfelds und eine sinkende Varianz der Blickpositionen zeigen z. B. Metz (2009) und Velichkovsky et al. (2003). Die Workload, die eine Aufgabe hervorruft, hat also einen messbaren Einfluss auf Scanning Techniken.

Das Problem beim Zweitaufgabenparadigma ist die **Interferenz mit der Erstaufgabe** (Vidulich, 2002), da die Zweitaufgabe bei hoher ökologischer Validität und hohen Anforderungen der Erstaufgabe per se in den Hintergrund tritt und die Varianz der Leistung unpräzise reflektiert wird (DeWaard, 1996). DeWaard (ebd., S. 35) weist darauf hin, dass „[t]he use of secondary tasks in applied environments is more complex than in laboratory experiments, and for this reason caution is required“. Aus diesem Grund wurde die Durchführung einer Zweitaufgabe in dieser Studie nicht favorisiert. Folgt man dem Autor (ebd.), ist es nötig, kombinierte Methoden für die Workload-Messung heranzuziehen, um valide Aussagen zu treffen. Gleichzeitig geben Svensson et al. (1999, S. 5) zu bedenken, „that performance measures are important in their own right and it is confusing to use them as aspects of PMWL [Pilot Mental Workload]“. Deutlich wird, dass hinsichtlich einer **Trennung von Workload- und Leistungsmaßen Dissens besteht** und eine Methodenkombination somit einen geeigneten Kompromiss darstellen dürfte.

Diesen Gedanken aufgreifend bilden **physiologische Workload-Indikatoren** eine geeignete Kombinationsmöglichkeit, deren Validität für die Workload-Messung evident ist (empirische Befundlage s. Roscoe, 1992). Die Messung der Herzrate (HR), Herzratenvariabilität (HRV), elektrodermalen (EDA) oder hirnelektrischen Aktivität (EEG) zählen bei der Messung von Workload zu den objektiven Verfahren (vgl. Roscoe, 1992; Svensson et al., 1999; einen ausführlichen Überblick liefert DeWaard, 1996, S. 37ff). Den physiologischen Indikatoren liegt die Annahme zugrunde, dass bei steigenden Aufgabenanforderungen (schnelle Informationsverarbeitung, sammeln und filtern von Reizen, schnelle Entscheidungsauswahl und Handlung) ein messbarer **Anstieg der neurologischen Aktivität bzw. der Erregung** zu verzeichnen ist, welcher mit entsprechenden Methoden abgebildet werden kann (Roscoe, 1992). Da **HR** ($r = .81$ bei $N = 40$) und **HRV** ($r = |.65|$ bei $N = 10$)

signifikant und in befriedigendem bis hohem Maß **mit dem NASA-TLX korrelieren** (Miyake, 2001; Lee & Liu, 2003; Lin, Imamiya, & Mao, 2008), kann davon ausgegangen werden, dass das weniger aufwändige Verfahren (hier: NASA-TLX) eine vergleichbar valide aber ökonomischere Messung der Workload darstellt. Aus diesem Grund wurde auf den Einsatz dieser physiologischen Methoden zur Bestimmung der Workload verzichtet. Für die Validität der **EDA** als Indikator liegen in der Abteilung Ergonomie des Flugmedizinischen Instituts der Luftwaffe (2012) eigene Erkenntnisse aus Studien im EC 135 Simulator vor. Hierbei ergab sich ein eher unzureichendes Potenzial des Verfahrens zur Messung der subjektiven Workload aufgrund sehr **hoher interindividueller Schwankungen und Spontanfluktuationen**. Auf Basis dieser Erfahrungen wurde auch die EDA aus der vorliegenden Studie ausgeschlossen. Daneben wurde auf die Erstellung eines **EEG** zur Workload-Messung verzichtet, weil die hirnelektrische Aktivität stark mit **Augenbewegungsartefakten** belastet sein kann (Joyce, Gorodnitsky, & Kutas, 2004). Außerdem ist die Elektroenzephalografie **technisch sehr aufwändig** und hätte einen weiteren (neben der Eye Tracking Apparatur) externen Monitor sowie Elektroden am Kopf der Probanden unter dem Pilotenhelm bedeutet. Für eine Initialanalyse der Scanning Techniken von Hubschrauberpiloten hätte dies den Rahmen des Machbaren und für die Probanden Akzeptablen sicherlich überschritten.

2.2.9 Aufgabenanforderungen: Komplexität vs. Schwierigkeit einer Mission

Wie bereits erwähnt, wird die mentale Workload neben dem Ressourcenportfolio des Operateurs durch **Aufgabenanforderungen (Demand)** moduliert: „*Demand is determined by the goal that has to be attained by means of task performance [e.g. 'the aircraft should land safely', S. 17], and is, once the goal has been set, external and independent of the individual. Load or Workload will be used to describe the effect the demand has on the operator [...]*“ (DeWaard, 1996, S. 15). Buld (2000) geht davon aus, dass Anforderungen, die sich aus der fliegerischen Tätigkeit ergeben, vorwiegend in den Bereich des Problemlösens fallen. Problemlöseszenarien sind Situationen, „*die vom Menschen zunächst nicht vollständig verstanden werden, aber von einem gegenwärtigen Ist-Zustand in einen angestrebten Zielzustand [überführt werden sollen]*“ (Johannsen, 1993, S. 323, zitiert nach Buld, S. 71f). Johannsen (ebd.) sieht das Durchführen von Fehlerdiagnosen und Planungstätigkeiten als zentrale Problemlöseinhalte bei der Mensch-Maschine-Interaktion. Dörner (1993, zitiert nach Buld, S. 72) beschreibt Problemmerkmale wie folgt:

- Komplexität: bei der Lösung muss eine Vielzahl untereinander abhängiger Variablen berücksichtigt werden,

- Vernetzung: die Variablen interagieren untereinander,
- Dynamik: die Problemstellung bzw. der Variablenkomplex entwickelt sich ständig weiter,
- Intransparenz: bestimmte Variablen bzw. Informationen sind dem Operateur nicht bekannt und/oder nicht zugänglich.

Der Umgang mit Anforderungen, die Problemlösetätigkeiten notwendig machen, lässt sich in die Schritte Situationsanalyse, Zielformulierung, Lösungssuche und Auswahl einer Lösung gliedern (ebd; vgl. Buld, 2000). Folgerichtig gilt: Je komplexer, vernetzter, dynamischer und intransparenter eine Aufgabe ist, desto höher ist die mentale Workload, da die Schritte des Problemlösens mehr Ressourcen erfordern (DeWaard, 1996; Buld, 2000). Bezogen auf die Situationsanalyse könnte sich dies beispielsweise in einer steigenden Blickdauer auf visuell komplexe Situations-elemente niederschlagen. Was aber bedeutet der Begriff **Komplexität**, (wie) lässt er sich operationalisieren und inwiefern ist er von Schwierigkeit abgrenzbar?

Ist Komplexität mit einem Wert beschreibbar?

Im Rahmen eines Versuchs Komplexität zu operationalisieren, führten Denning et al. (2003) eine Studie zur Komplexitätsindizierung von Kampfmissionen für F15 und F16 Waffensysteme durch. Bei Experteninterviews (Fluglehrer) stellte sich allerdings die “[...] *inability of subject matter experts to agree on valuation and combination of complexity into a single score for a scenario*” heraus (ebd., S. 19). Daher wurde die Komplexität der Luftkampfszenarien auf einer relativ allgemeinen, einer individuellen (bezogen auf Pilot), einer teambezogenen (bezogen auf zwei Luftfahrzeuge) und einer feindbezogenen Dimension beschrieben (S. 4ff):

→ allgemeine Komplexität: “[...] *task count, load, and pace was considered in the context of the skill set of both types of intercept to establish the lowest level of complexity to achieve a kill objective*”,

→ individuelle Komplexität: “[...] *an easily detected and discerned target, no counter-maneuver or awareness of the fighter by the target, no weapon employment restrictions, and no detractors to situation awareness development*”,

→ teambezogene Komplexität (zwei Singleseater): “[...] *additional mental and physical tasking of individuals within [two aircraft operating as a team], essential communications requirements between team members, and limitations imposed on available resources posed by additional stressors in or near the fight space*”,

→ feindbezogene Komplexität: “[...] *number of aircraft, detection range, sort ability at range, weapons carried, awareness of fighter attack, counter-maneuver, countermeasures, and shootable merit based on rules of engagement*”.

Das Beispiel soll verdeutlichen, dass sich Komplexität kaum mit einem Einzelwert niederschlägt, sondern eher auf verschiedenen Dimensionen erzeugt wird. Bei Denning et al. (ebd., S. 4ff) spielen vor allem numerische (“*task count*”, “*number of aircraft*”) und relationale Aspekte (“*no counter-maneuver or awareness of the fighter by the target*”) eine Rolle. Die Autoren schließen, dass die Definition und Operationalisierung von Anforderungen mit der detaillierten Auseinandersetzung mit Expertenmeinungen einhergehen muss, um valide Ergebnisse zu erhalten.

Die Termini **Schwierigkeit**, **Komplexität** oder **Kompliziertheit** werden im Alltag aber auch im Bereich Ergonomie teilweise synonym verwendet. Die Begriffswolke, die bei näherem Nachdenken hieraus entsteht, wird noch vergrößert, wenn man bedenkt, dass es verschiedene Schwierigkeits- oder Komplexitätsbereiche gibt (z. B. Systemkomplexität vs. Aufgabenkomplexität oder Aufgabenschwierigkeit vs. Handlungsschwierigkeit): „*One of the first obstacles to establishing valid observations was clearing up conceptual fog regarding difficult versus complex. Observing in other disciplines that complex items are usually not difficult for a master, while simple tasks may prove difficult to a novice, it was clear that objective complexity needed to be employed to eliminate sensitivity to proficiency and/or automaticity if these are indeed the targets to be evaluated in training*” (Denning et al., 2003, S. 14). Aus dem Zitat geht übereinstimmend mit DeWaard (1996, S. 15) hervor, dass sich die Begriffe auf den Polen objektiv vs. subjektiv unterscheiden lassen, wobei die subjektive Komponente vor allem in Abhängigkeit der Erfahrung variiert.

Nach umfangreicher Literaturanalyse kommt Xing (2004; Xing & Manning, 2005) in Konkordanz zu den Befunden von Denning et al. (2003) zu dem Schluss, dass sich **Aufgabenkomplexität** auf drei Ebenen der menschlichen Informationsverarbeitung (Wahrnehmung, Kognition, Handlung) definieren lässt. Muss beispielsweise bei schwer zu beurteilenden Landeverhältnissen (Hanglage, kleine Landefläche) eine große Menge an Information aus den Bordinstrumenten entnommen und gleichzeitig die Umgebung gescannt und auf Referenz- oder Hindernispunkte geprüft werden, liegt eine hohe Wahrnehmungskomplexität vor. Dies gilt insbesondere, weil die Informationen aus verschiedenen Quellen (Cockpit vs. Umgebung) stammen und in variierender Beziehung stehen (z. B. Höhe vs. Anflugwinkel bei Landung über Gewässern). Um ein erfolgreiches Landemanöver vorbereiten und durchführen zu können, muss der Pilot verschiedene Informationen kombinieren, deren Folgen abschätzen und sich letztendlich für die folgerichtigste entscheiden (Kognitionskomplexität; z. B. Absetzen vs. Durchstarten). Die Vielzahl an Steuerin-

strumenten, Bordelektronik und deren technisches Zusammenspiel verleihen der Luftfahrzeugführung eine hohe Handlungskomplexität. Was in der Betrachtung von Xing keine Erwähnung findet, sind die Konsequenzen einer Handlung, die zwar nicht direkt die Informationskomplexität kontaminieren, aber die Wahrnehmung beeinflussen können (vgl. Wickens, 2009). Für den Operateur kann die Gesamtsituation bei schwerwiegenden Handlungskonsequenzen in ihrer Komplexität gesteigert werden (z. B. in Luftkampfscenarien oder bei der Personenrettung).

Denning et al. (2003, S. 14) sehen den Zusammenhang zu Schwierigkeit wie folgt: *“Difficulty resides more in the cognitive domain as a measure of how an operator’s proficiency in an event measures up in handling the realized complexity of the event”*. Hier deutet sich erneut an, dass Komplexität auf den oben dargestellten Dimensionen ein reines Aufgabenmerkmal – oder im Sinne DeWaards (1996) ein Korrelat der Demands – ist, während Schwierigkeiten bei der Aufgabenbearbeitung vornehmlich durch Informationsverarbeitungsprozesse des Operateurs moduliert werden. Denning et al. (ebd., S. 14) sehen ergänzend zwei Arten von Schwierigkeit: *“When analyzing, the internal structure of the [complexity] indices [...], we found that the markers could be ordered in a sequence. One end represents items of perceptual content, e.g. difficulties in surveying the symbolic representations [on a display], and the other end represents items of cognitive content, e.g. difficulties in integrating information and make decisions. [...] It is interesting to note, that the workload measures are more related to the cognitive aspects [...] than to the perceptual aspects [...]”*. Mit dem letzten Teil des Zitats wird impliziert, dass Diskrepanzen zwischen Workload- (z. B. Selbstreport) und perzeptuellen Messverfahren (z. B. Eye Tracking) nicht auszuschließen sind, da Erstere eher kognitive, Letztere eher wahrnehmungsbezogene Anforderungen abbilden.

Theofanou (2002, S. 15f; in Anlehnung an DeWaard, 1996) unternimmt den Versuch, die Begriffe Schwierigkeit und Komplexität zusammenfassend zu differenzieren: *„[Komplexität] steigt mit der Anzahl der Verarbeitungsstufen die benötigt werden, um eine Aufgabe zu bearbeiten. Die Schwierigkeit [...] bezieht sich auf die ‘Processing effort (amount of resources) that is required by the individual for task Performance’“*. Dies impliziert, dass hohe Aufgabenkomplexität mit ineinandergreifenden und kaum voneinander abgrenzbaren Aufgabenanforderungen verbunden ist, die im experimentellen Kontext nur schwer zu operationalisieren sind. Im Gegensatz dazu erfordert hohe **Aufgabenschwierigkeit** (hier: Processing Effort) offenbar ein hohes Maß an Geschick, Umsicht und Erfahrung in speziellen zur Aufgabenerfüllung notwendigen Tätigkeiten (in diesem Sinne also Teilziele, vgl. DeWaard, 1996), die sich mithilfe der Erfahrung eines Experten identifizieren lassen dürften. Ein Vergleich der Begriffe Komplexität und Schwierigkeit ergibt sich

abschließend anhand der Analogie von Belastung vs. Beanspruchung: Während Belastung der objektive Einfluss auf ein Individuum ist (vergleichbar mit objektiver Komplexität), wird unter Beanspruchung das subjektive Empfinden unter Berücksichtigung der verfügbaren Ressourcen des Operators verstanden (vergleichbar mit subjektiver Schwierigkeit). DeWaard (1996, S. 21) konkludiert: „*Task difficulty depends upon task complexity, the operator’s possibilities (i.e. capacity), his or her state and the applied strategy. Finally mental workload [...] is determined directly by task difficulty.*“ Aufbauend auf den Erkenntnissen von Denning et al. (2003) werden Aufgabenanforderungen bei Missions in dieser Arbeit mithilfe von Expertenmeinungen operationalisiert (s. Abschnitt 3.3.1). Welche Befunde liegen aber bisher zu Anforderungen im Zusammenhang mit Scanning Techniken vor?

Aufgabenanforderungen und Befunde zu Scanning Techniken von Piloten

Colvin et al. (2003) untersuchten Blickwege (Scan Paths) von Piloten bei Variation der Aufgabenschwierigkeit (N = 5). Neben normalen Überflugphasen wurden Phasen mit hoher Luftverkehrsdichte untersucht. Die Einzelfallanalyse zeigt, dass bei steigenden Anforderungen eine Konzentration der Blicke auf zentrale Displays und Instrumente erfolgt, um damit das Leistungsniveau (Höhe, Kurs, Geschwindigkeit) halten zu können. Im Gegensatz dazu werden Seitendisplays weniger häufig fixiert. Der Vergleich zwischen den Blickwegen der Piloten zeigt, dass diese keinen über alle generalisierbaren Blickweg aufweisen.

Blickkonzentrationen auf die Instrumentierung untersuchten auch Thomas und Wickens (2004), indem sie Kontrolldisplays im Cockpit variierten (N = 8). Die Displays wurden einmal mit und einmal ohne grafisches Raster, welches wie ein optischer Tunnel wirkte, ausgestattet. Gemessen wurden Scanning Techniken und ob die Piloten unerwartete andere Flugzeuge im Luftraum entdecken. Es zeigt sich in dieser Studie, dass Non-Detectors beide Displays mehr fixierten als Detectors und weniger Blicke nach außen tätigten. Detectors hingegen zeigten eine gleichmäßigere Verteilung ihrer Blicke sowohl auf die Displays als auch nach außen. Erwartungsgemäß wirkte sich der Tunnel des einen Displays negativ auf die Entdeckungsleistung anderer Flugzeuge aus, da er Visual Tunneling begünstigte. Auf Visual Tunneling wurde allerdings auch in dieser Studie nicht anhand der Blickdauer geschlossen, sondern mithilfe von Vergleichen der Blickhäufigkeit.

Während die beiden genannten Studien bei steigenden Anforderungen Visual Tunneling Effekte nachweisen können, impliziert die Arbeit von DiNocera, Camilli und Terenzi (2007) den reziproken Trend. Die Autoren erhoben die subjektive Workload, die Leistung und Scanning Techniken in Abhängigkeit der Flugphase (Start – Überflug – Landung) bei Polizeipiloten (N = 10). Die Ergebnisse zeigen: Je höher die Workload war (Start- und Landephase), desto zufälliger bzw. ungeziel-

ter sind die Blickpunkte der Probanden im simulierten Cockpit; sie zeigen eine kürzere Blickdauer und längere Sakkaden (Visual Scanning Randomness). Die Autoren mutmaßen, dass durch zufälligeres Umherblicken die Informationsaufnahme optimiert werden soll, um auch in Phasen mit hoher Workload nichts zu verpassen (vgl. hierzu auch Graeber & Andre, 1999, NASA Ames Research Center). Außerdem vergleichen sie die Befunde mit denen zur Betrachtung von Bildern: Am Anfang der Bildbetrachtung (hohe Workload, da Bild unbekannt) seien Fixationen kürzer und Sakkaden länger um damit zunächst strukturelle Informationen aufzunehmen. Wenn der Beobachter an das Bild adaptiere (Workload sinkt wieder, da Bild bekannt), würden Fixationen länger und Sakkaden kürzer werden, da semantische Informationen aufgenommen werden (vgl. DiNocera et al., 2007).

Svensson, Angelborg-Thanderz, Sjoberg und Olsson (1997, Swedish Air Force) untersuchten das Verhältnis von Head Up und Head Down Blicken bei Kampfpiloten (N = 18). Sie variierten die Flugaufgabe (Navigation vs. Luftraumüberwachung) sowie deren Anforderung (Anzahl feindlicher Flugzeuge) und erhoben die Leistung (Höhenabweichung), subjektive Workload und Scanning Techniken bei Kampfpiloten. Es zeigt sich, dass bei steigender Beanspruchung ein häufigerer Blickwechsel zwischen Head Up und Head Down stattfindet, das heißt Blicke in diese Regionen kürzer, Sakkaden jedoch länger und häufiger werden.

Wie der empirische Exkurs verdeutlicht, kann kein einheitlicher Rückschluss von steigenden Aufgabenanforderungen auf Scanning Techniken von Piloten gezogen werden. Auf der einen Seite steigt die Tendenz zu Visual Tunneling, auf der anderen wird von zufälligem Umherblicken berichtet. Da die einzelnen Operationalisierungen der Aufgabenanforderungen nicht vergleichbar sind, bleibt die Frage nach Generalisierbarkeit zunächst offen. Denkbar ist an dieser Stelle der Einfluss eines weiteren Faktors auf die Varianz der Scanning Techniken: Die Flugerfahrung und der Abruf von gelernten Strategien aus dem Langzeitgedächtnis.

2.2.10 Erfahrung und Langzeitgedächtnis

Das Ressourcenmanagement und die wahrgenommenen Aufgabenanforderungen werden neben den bisher diskutierten Faktoren durch den **Erfahrungsstand** und damit das **Langzeitgedächtnis** (s. Abbildung 22) des Operators beeinflusst (vgl. auch Herczeg, 2005, S. 83ff). Während unerfahrene Operateure noch keinerlei Automatismen bei der Bearbeitung einer Aufgabe entwickelt haben, können Experten relativ leicht und flexibel neue und schwierige Aufgabenelemente lösen und das System an sich assimilieren: *“Learned skill and factual knowledge, including knowledge of strategies, seem to be the dominant source of performance differences between experts and novices [...]”* (Svensson et al., 1999, S. 11).

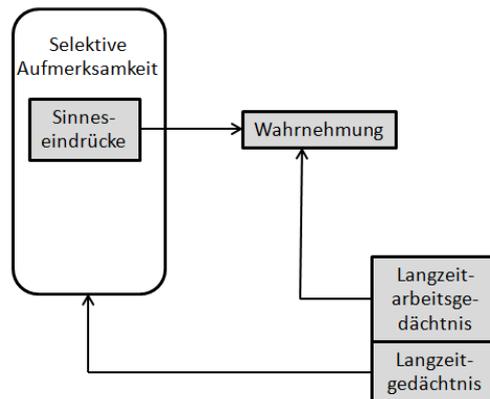


Abbildung 22: Navigation im Modell von Wickens (2009):
Langzeitgedächtnis und Langzeitarbeitsgedächtnis.

Die **Überlegenheit** von Experten gegenüber Novizen erklärt O´Hare (2002) mithilfe einer auf das Langzeitgedächtnis bezogenen Strategie: Experten versuchen erfahrungsgelenkt, situationsrelevante Reize zu identifizieren, die zur Lösung einer Aufgabe beitragen können (auch: **Langzeitarbeitsgedächtnis**). Irrelevante Reize werden ausgeblendet, sodass in einer schwierigen Situation mehr kognitive Ressourcen zur Verfügung stehen als bei Novizen. Tsang (2002) diskutiert Wirkfaktoren, die den Erfahrungsstand von Piloten modulieren. Hierunter fällt das enorme Ausmaß intendierter Übung, welche sich aufgrund von Metakognitions- und Fehlerkorrekturtätigkeiten von rein spielerischer Interaktion unterscheidet. Darüber hinaus geht steigende Erfahrung bei der Flugaufgabe mit wachsendem **domänenspezifischen Wissen** einher (Anderson, 1989; Tsang, 2002); also mit Lösungen für spezifische Problemstellungen (O´Hare, 2002). Die Vorteile hoher Expertise lassen sich wie folgt zusammenfassen (vgl. Tsang, 2002, S. 518; O´Hare, 2002; S. 226):

- Gut strukturierte Wissensbasis für leichten Informationsabruf,
- tiefgründige Problemrepräsentation – auch bei nicht-transparenten Problemen sowie Skepsis gegenüber dem „normalen“ Zustand des Luftfahrzeugs,
- Fähigkeit, relevante Muster zu erkennen, Beziehungen zwischen Informationen herzustellen, die Wahrnehmung zu aktualisieren und den zukünftigen Verlauf einer Situation vorherzusehen (Situationsbewusstsein),
- adäquate Handlungsauswahl aus dem Langzeitspeicher bei neuartigen Problemen bzw. automatisiertes Handeln in bekannten Situationen,
- Fähigkeit zur Metakognition über die eigene Leistung und den Stand der eigenen Ressourcen sowie der Ressourcen von Team-Mitgliedern (Perspektivenübernahme) und
- Selbstvertrauen sowie die Motivation zu lernen.

Wie aber funktioniert das Anlegen domänenspezifischen Wissens, sodass Experten von diesen Vorteilen profitieren können? Tsang (2002) leitet dies aus der Betrachtung her, dass das Langzeitgedächtnis unabhängig von Kurz- und Arbeitsgedächtnis einzuordnen ist (Ericsson & Delaney, 1998). Während Kurz- und Arbeitsgedächtnis in ihrer Kapazität als begrenzt gelten (7 ± 2 Informationseinheiten) und dies sowohl für Experten als auch Novizen gilt, ist das Langzeitgedächtnis hinsichtlich seiner Speicherkapazität theoretisch unbegrenzt. Die Güte des Abrufs aus dem Langzeitspeicher ist also kein quantitativer, kapazitätsbezogener Aspekt, sondern eine qualitative Frage der Indizierung von Information. Mit anderen Worten geht es um den Schlüssel oder Code, der Speicherinhalten zugewiesen wurde und der sich zwischen Erfahrungsgruppen unterscheidet. Darauf aufbauend zitiert Tsang (2002, S. 521): „*Ericsson and Charness (1997) pointed out that experts acquire memory skill to meet specific demands of encoding and accessibility in specific activities. For this reason, their skill is unlikely to transfer from one domain to another.*“ Kombiniert man den höheren Anteil an domänenspezifischem Wissen von erfahrenen Piloten mit der Tatsache, dass sie von einem größeren funktionellen visuellen Feld profitieren (s. Abschnitt 2.2.6), schneller relevante Muster erkennen und adäquate Handlungsentscheidungen treffen, ist davon auszugehen, dass sie in Flugaufgaben (Domäne bleibt gleich) mit hohen Anforderungen zu von Novizen klar unterscheidbaren Scanning Techniken tendieren. Dies wird durch einige empirische Arbeiten bereits angedeutet.

Flugerfahrung und Befunde zu Scanning Techniken von Piloten

Kasarskis et al. (2001, S. 2) übertragen den Einfluss der peripheren Wahrnehmung auf Scanning Techniken erfahrener Operateure vom Fahrkontext auf Piloten: „*Mourant and Rockwell (1972) found that expert drivers looked farther down the road and gained more lane information from their peripheral vision. This allowed experts to stay in their lane better than novices. In contrast, novices demonstrated a behavior in which they glanced slightly ahead and to the right of the car gathering lane position information directly from seeing the distance from the curb. Not only did this prevent them from looking forward, but it also hindered their ability to stay within their lane. Wikman, Nieminin & Summala (1998) also found that novices' in-car dwells hindered their ability to maintain lateral stability. Perhaps experts were able to gain peripheral information about the outside position of the car, even when their fixations were inside. In support of this hypothesis, Fox, Merwin, Marsh, McConkie, and Kramer (1996) found that performance for expert pilots suffered more than for students when flight relevant information was removed from the instruments in the pilots' peripheral vision. Therefore, they concluded that experts are more likely than novices to use peripheral vision and*

hence, process a broader range of visual cues, thereby reducing the need to make a conscious fixation on an instrument and allowing the experts to vary their scanning pattern while still obtaining the needed information."

Hinsichtlich der Flexibilität von Experten in gleichbleibenden Domänen untersuchten Dixon, Rojas, Krueger und Simcik (1990, US Air Force) Scanning Techniken und die Leistung bei unterschiedlich erfahrenen C 130 Piloten (N = 12) und variierten die Größe des Sichtfeldes (Field of View; FOV) im Flugsimulator. Nachweislich adaptieren erfahrene Piloten im Gegensatz zu Novizen effizienter an ein kleineres Sichtfeld, um die Flugparameter zu halten. Die Strategie der Experten bestand daraus, bei kleinerem FOV - schneller als Novizen - weniger OTW, aber mehr auf die Instrumente zu blicken.

In Bezug auf IFR-Flüge fanden Bellenkes, Wickens, and Kramer (1997) verschiedene Scanning Techniken von Novizen und Experten. Die Blickdauer der Experten war insgesamt geringer und Fixationen auf die Instrumente waren höher frequentiert im Vergleich zu Novizen. Dadurch waren erfahrene Zivilpiloten eher in der Lage flexibel auf variierende Aufgabenanforderungen (hier: Flugphasen) zu reagieren.

Ein ähnliches Ergebnis erzielten Kasarskis et al. (2001) für Landemanöver von Zivilpiloten unter VFR-Bedingungen. Hier zeigten Experten mehr relevante Fixationen auf Zielregionen oder Instrumente (Target Fixations), verfügten damit über die gezielteren Scanning Techniken und platzierten ihre Aufmerksamkeit effektiver als unerfahrene Piloten.

Sullivan et al. (2011) untersuchten Scanning Techniken von unterschiedlich erfahrenen Hubschrauberpiloten (N = 12) in einem fest stationierten Simulator mit einer Navigationsaufgabe. Im Ergebnis zeigt sich, dass die Leistung (hier: Abweichung von einem optimalen Weg) nicht durch die Flugerfahrung vorhersagbar ist, Scanning Techniken jedoch korrelieren mit der Expertise: Je höher die Flugerfahrung ist, desto geringer ist die Blickdauer (M = 231 ms) und desto häufiger sind Blickwechsel (M = 123) zwischen Außenwelt und der Karte im Cockpit. Unerfahrenere Piloten blicken in dieser Studie entgegen den Erwartungen bei einer Navigationsaufgabe mehr OTW (M = 57.7 %) als Experten. Auch hier verfügen erfahrene Piloten also über die effizienteren Scanning Techniken.

Das Prinzip der strategischen Informationsakquise durch Experten zeigt sich auch in der Raumfahrt: Matessa und Remington (2005) modellieren das Verhalten von Astronauten beim Fehlermanagement, indem sie die Verhaltenssequenzen im Space Shuttle hierarchisch zerlegen (Alarm, Identifizierung, Korrekturprozedur auswählen, Korrektur durchführen, Verifizierung). Ausgegangen wurde von einer modellbasierten Gesamtblickdauer von 5170 ms für die Bearbeitung einer Feh-

lermeldung. Die Stichprobe (N = 10) bestand jeweils zur Hälfte aus trainierten Astronauten (Experten) und zivilen Airline-Piloten (hier „Novizen“), die jeweils einen plötzlichen Fehler während eines simulierten Space Shuttle Fluges bearbeiten sollten. Die Autoren legten verschiedene AOIs fest und zeichneten die Blickdauer auf. Es gab insgesamt fünf AOIs: 1. Message (hier erschien die Fehlermeldung), 2. Data (hier sollte der Fehler identifiziert werden), 3. Keyboard (hier wurde operiert), 4. Procedures (hier wurden Anleitungen für Handlungen gegeben) und 5. Switch (hier sollte ein Hebel umgelegt werden). Im Ergebnis zeigt sich: Novizen fixieren die Alarm- bzw. Message-Region, in der der Fehler gemeldet wurde, und die Procedures-Region länger als Experten. Im Gegensatz dazu fixieren die Experten das Data-Display länger als Novizen. Die Resultate lassen den Schluss zu, dass Novizen im Vergleich zu Experten länger und wiederholt plötzlich auftretende, unbekannte Informationen verarbeiten müssen, während erfahrene Astronauten schneller eine effektive Statusanalyse des Space Shuttle durchführen.

Fasst man die Erkenntnisse zusammen, kann nicht verallgemeinert werden, dass eine Erfahrungsgruppe kürzer oder länger bzw. häufiger oder seltener bestimmte Bereiche im Cockpit fixiert, sondern es ist eher davon auszugehen, dass Experten Blicke zur Informationsakquise mit einer gezielteren Intention einsetzen als Novizen. Flugerfahrung und Aufgabenanforderungen müssen also scheinbar als kombinierte Einflussgröße auf Scanning Techniken angenommen werden.

2.2.11 Situationsbewusstsein

Wie im Abschnitt 2.2.7 bereits anklung, wird als entscheidendes Konstrukt zur Beschreibung der Informationsrepräsentation und -verarbeitung von Piloten das Situationsbewusstsein diskutiert (s. Abbildung 23), da es einer adäquaten Entscheidungsfindung, der richtigen Handlungsauswahl und -ausführung vorausgeht. Nach Endsley (1995, S. 36; „Situational Awareness“; SA; s. auch, z. B. 1995; 1996; 2000; vgl. Vidulich, 2002; Svensson et al., 1999; Herczeg, 2005) beschreibt Situationsbewusstsein „*the perception of the elements in the environment within a volume of time and space (1), the comprehension of their meaning (2), and the projection of the status in the near future (3)*“. Die Stufen 1 bis 3 des Situationsbewusstseins werden sowohl durch Aufgaben-/Systemfaktoren (z. B. Schnittstellengestaltung oder Komplexität) als auch durch Faktoren des Operateurs (z. B. Erfahrung) determiniert. Die einzelnen Faktoren sind in Abbildung 24 dargestellt.

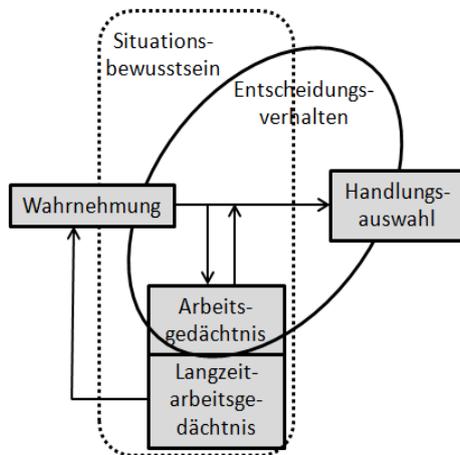


Abbildung 23: Navigation im Modell von Wickens (2009): Situationsbewusstsein.

Die Autorin fasst zusammen, dass Situationsbewusstsein auch mit *“knowing what is going on”* (Endsley, 2000, S. 2) umschrieben werden kann. Unterschieden werden außerdem verschiedene Arten von SA (Endsley, 1996): z. B. geografisches, räumliches und zeitliches, umweltbezogenes und systembezogenes Situationsbewusstsein. Bei normalen Arbeitsabläufen spielt SA als Ressource von Piloten eine untergeordnete Rolle; es tritt erst in den Vordergrund, wenn zielgerichtetes und effizientes Handeln in unerwarteten Situationen gefordert ist (Wickens, 1996).

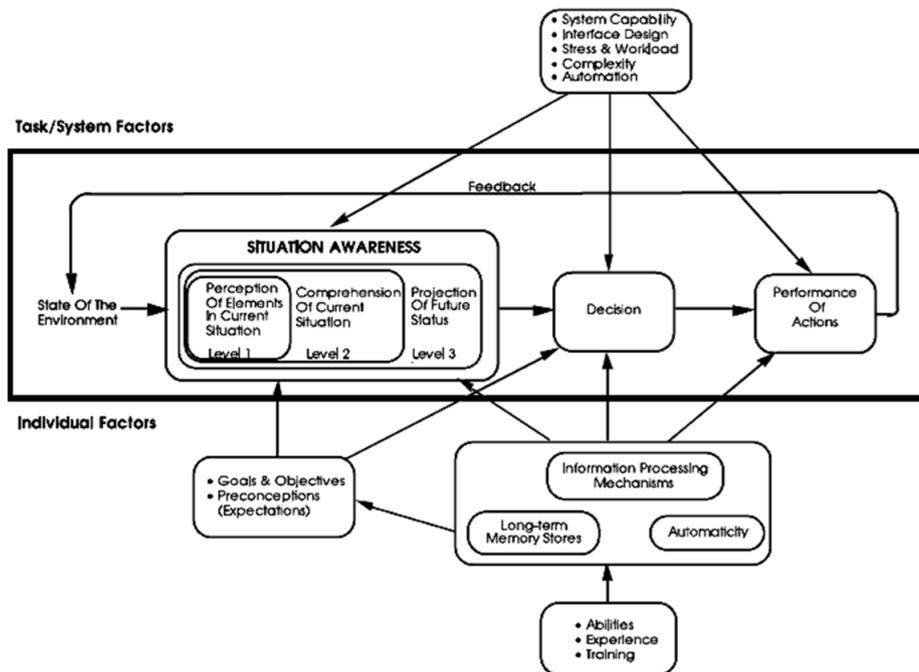


Abbildung 24: Modell des Situationsbewusstseins (übernommen aus Endsley, 1995).

Folgt man Vidulich (2002, S. 177) ist Situationsbewusstsein weniger bezogen auf die quantitative Ressourcenverteilung, sondern *“[...] situation awareness is more concerned with the quality of the information apprehended by the pilot”*. Basierend auf Endsleys (2000) Ausgangspunkt eines Informationsdefizits (Information Gap) ist es die zentrale Herausforderung für Piloten aus einer komplexen Ansammlung von Information (Data Produced) die relevanten Parameter (Information Needed) herauszufiltern, um ein optimales Situationsmodell zu erhalten. In diesem Sinne müsste die Kurzfassung der Definition statt *„knowing what is going on“* eher *„knowing what is needed“* heißen. Hierzu hält Wickens (1992, zitiert nach Uhlarik & Comerford, 2002, S. 2) fest: *„It is the ability to mentally access relevant information about the evolving circumstances of a flight.“* Crane (1992, zitiert nach ebd., S. 2) sieht SA als die Fähigkeit, auf einem Expertenniveau zu operieren. Sind mehrere Operateure in demselben Handlungskontext aktiv, sollten sie zu einem geteilten mentalen Abbild der Situation gelangen.

Situationsbewusstsein bei Kampfmissionen

In einer Studie zur Missionsevaluation umschreiben Svensson et al. (1999, S. 14) das Situationsbewusstsein mit Situationskenntnis (Situational Cognizance; SC) als latente Variable, die die Leistung bei einer Kampfmission beeinflusst. Für die Autoren manifestiert sich das Konstrukt SC in den folgenden sieben Indizes, mit denen Piloten befragt wurden:

- *“To what extent could you estimate the flight paths,*
- *was the course of event as expected,*
- *could you predict the mission course of events,*
- *was the cooperation within the group functioning well,*
- *did you have ‘mental lead’ with respect to the air defense task,*
- *did you recognize the course of events, and*
- *were you in control of the situation?”*

Das Beispiel zeigt, dass SA nicht nur synonym umschrieben werden kann, sondern offenbar auch je nach operationellen Erfordernissen des Kontextes heterogen gemessen wird (Seitz & Häcker, 2008, zitiert nach Herczeg & Stein, 2012).

Der **Zusammenhang**, in dem SA zu den anderen Konstrukten steht, wird kontrovers diskutiert. Während Endsley (1995) SA als Vorstufe der Entscheidungsfindung, aber separat von Leistung betrachtet, differiert die Meinung von Svensson et al. (1999, S. 9) hierzu: *“According to our opinion and experience, SA is often so closely related to the pilot’s performance that it is more logical to consider it a part of performance”*. Mit dieser Ansicht geht auch Vidulich (2002, S. 137) kon-

form, denn “[...], a review of situation awareness measurement in studies of interface manipulations found that the manipulations of interface design undertaken to improve situation awareness tended to improve task performance”. Noch unüberschaubarer ist die Diskussion um den Zusammenhang zwischen SA und Workload. Vidulich (2002, S. 117) differenziert beides hinsichtlich des Globalitätsgrades: „[...], mental workload is the cost placed on the pilot by performing the necessary mental processing to accomplish the mission, and situation awareness is the momentary understanding of the current situation and its implications”, aber auch hinsichtlich des kognitiven Korrelates (S. 122): “[...] workload can be seen as primarily an attentional phenomenon, whereas situation awareness is perceived to be more strongly associated with memory”. Auch Endsley (1993) betrachtet Workload und SA als unabhängige Konstrukte, da es ihr konzeptuell unlogisch erscheint, die Ausprägung der einen Variablen aus der anderen vorherzusagen. Vidulich (2002) bestätigte dies mithilfe einer Metaanalyse von 15 Studien, allerdings finden einzelne empirische Arbeiten einen linearen Zusammenhang: Je höher die Workload ist, desto eher sinkt das Situationsbewusstsein (z. B. Svensson et al., 1999). Denkbar ist hier, dass Inkonsistenzen auftreten, weil die Konstrukte auf verschiedene Arten operationalisiert werden oder weil im Rahmen konfirmatorischer Hypothesen nach linearen Zusammenhängen anstelle von nicht-linearen (z. B. quadratisch: SA ist am höchsten bei mittlerer Workload) gesucht wird.

Wie aber kann SA überhaupt **gemessen** werden? „Most real world evaluations of situation awareness do not attempt to directly measure situation awareness, but implicitly infer changes in situation awareness from changes in overall performance [...]” (Vidulich, 2002, S. 119). Neben diesem induktiven Messansatz, beschreibt der Autor ein weiteres prototypisches Verfahren, den Gedächtnistest **SAGAT** (Situation Awareness Global Assessment Technique; Endsley, 1988), welcher aus mehreren Schritten besteht. Zunächst werden mithilfe von Expertenbefragungen Items gesammelt, die für die Bildung von SA relevant sind (z. B. Flughöhe, Anzahl feindlicher Luftfahrzeuge, taktischer Status), danach wird beispielsweise ein Simulatorflug durchgeführt, welcher in zufälligen Zeitintervallen gestoppt wird (Freezing) um eine Auswahl der Items abzufragen. Je nach Anzahl richtig beschriebener Items durch den Piloten errechnet sich ein SA-Wert kumulativ. Die SAGAT ist aufgrund der intrusiven, wenig objektiven Durchführungserfordernisse, mangelnder Sensitivität (Hughes, Hassoun, Ward, & Rueb, 1990) und theoretischer Fundierung **kritisch** zu bewerten und bedarf Forschungsbemühungen zur Weiterentwicklung (Vidulich, 2002).

In summa vereint Endsleys Modell die wichtigsten kognitionspsychologischen Konzepte, die bei der HMI eine Rolle spielen, auf deskriptivem Niveau. Entspre-

chend stellt der Ansatz im Vergleich zum grundlegenden Modell der Informationsverarbeitung (Wickens, 2009) eine detailliertere Zusammenfassung der kognitiven Prozesse dar, die hier diskutiert wurden. Während sich SA mit einer klaren Augenscheinvalidität präsentiert, bleibt aber zu betonen, dass es deskriptiven Charakter hat und sich mit in der Kognitionspsychologie bereits hinlänglichen etablierten Konstrukten schmückt. Es ist fraglich, ob es sich dabei tatsächlich um eine distinkte Modellierung von kognitiven Prozessen im Cockpit handelt, oder eher um eine treffende Synonymisierung theoretischer Grundlagen der menschlichen Informationsverarbeitung (vgl. Sarter & Woods, 1991). Das SA-Konzept ist zudem noch nicht hinreichend differenziert, die Messmethoden können nicht als standardisiert oder validiert gelten und es herrscht ein **Defizit** an theoriegeleiteter Forschung auf diesem Gebiet (Seitz & Häcker, 2008, zitiert nach Herczeg & Stein, 2012). Aufgrund der Mängel wurde SA im empirischen Teil der Arbeit nicht näher fokussiert.

2.2.12 Schlussfolgerungen für die Untersuchung

In der vorliegenden Arbeit werden die Scanning Techniken von Hubschrauberführern bei simulierten Landemanövern analysiert. Vor diesem Hintergrund leiten sich aus der theoretischen Darstellung folgende Schlüsse ab:

1. Dem Grundlagenmodell von Wickens (2009) zufolge wird **Informationsverarbeitung** als sequenzieller Prozess von Aufmerksamkeit und sensorischer Wahrnehmung hin zur Handlungsexekution verstanden. Aufmerksamkeit und sensorische Wahrnehmung werden durch das Informationsmanagement im Arbeits- und Langzeitgedächtnis beeinflusst, sind die Vorstufe von Entscheidungen des Piloten (Aeronautical Decision Making) sowie von Handlungsauswahl und -exekution. Handlungen des Piloten können wiederum seine Aufmerksamkeit und Wahrnehmung im Sinne einer Rückkopplung beeinflussen.
2. Dem theoretischen Diskurs von Herczeg und Stein (2012) zufolge ist unter anderem von einem Einfluss von **affektiven, motivationalen und ermüdungsbezogenen Zuständen** auf die Informationsverarbeitung auszugehen. Einige Studien (Changizi & Hall, 2001; Kimble et al., 2010; Bradley et al., 2012) untermauern diese Annahmen für die späteren Ebenen der Interpretation von Informationen und Handlungsauswahl. Die Annahme eines Effektes dieser Faktoren auf die Ebene der Aufmerksamkeit und damit auf Scanning Techniken von Hubschrauberpiloten kann auf Basis recherchierter empirischer Befunde (Pannasch et al., 2008; Previc et al., 2009; Rohani, 2012) nicht als gesichert gelten und wird daher in dieser Arbeit nicht näher fokussiert. Ein Einfluss auf andere Indikatoren der Informationsverarbeitung (z. B. durch Motivbezogene Interpretation von Stimuli) ist nicht auszuschließen.

3. **Scanning Techniken** werden als messbare Indikatoren der Informationsverarbeitung von Hubschrauberpiloten, hier insbesondere auf der Ebene der Aufmerksamkeit und visuellen Informationsakquise, betrachtet. Den Aussagen von Fluglehrern der Heeresfliegerwaffenschule zufolge werden **Target Fixations** als eine auf den Hubschrauberkontext bezogene Scanning Technik verstanden. Zur Messung wird die Eye Tracking Methode verwendet, da sie derzeit das einzige adäquate Verfahren hierfür darstellt. Es wird davon ausgegangen, dass im Kontext der Hubschrauberführung für die Mehrzahl der Target Fixations die Eye-Mind-Hypothese (Just & Carpenter, 1980; vgl. z. B. Poole & Ball, 2000; Tsai et al., 2007; Metz, 2009; Geise, 2012) gilt, wonach der Blickort gleichzeitig der Ort der Aufmerksamkeit von Hubschrauberpiloten ist. Damit werden Interpretationsunschärfen in Kauf genommen, denn die Eye-Mind-Hypothese gilt nicht immer (Davidson & Parker, 2001; Anderson et al., 2004) und es kann damit nicht zwischen endogener und exogener sowie nicht direkt zwischen peripherer und fovealer visueller Wahrnehmung unterschieden werden. Entsprechend werden die gemessenen Target Fixations der Empfehlung von Geise (2012) folgend im Rahmen des Experimentalparadigmas unter den Aspekten Flugerfahrung und Missionsanforderungen interpretiert.
4. Die **Leistung (Performance)** während einer Mission wird ebenfalls als messbarer Indikator der Informationsverarbeitung von Hubschrauberpiloten, hier insbesondere als Gütekriterium des Informationsmanagements, betrachtet. Den Erkenntnissen von DeWaard (1996), Buld (2000) oder Kasarskis et al. (2001) zufolge reflektiert die Varianz in der fliegerischen Leistung Erfahrungsunterschiede zwischen Piloten. Die Leistungsmessung erfolgt üblicherweise über Parameter der Hauptaufgabe (z. B. Abweichung von einem Sollwert, Reaktionszeit) oder über die Güte der Erfüllung einer parallelen Sekundäraufgabe. Das Sekundäraufgabenparadigma wurde hier nicht fokussiert, weil Zweitaufgaben derart mit der Hauptaufgabe interferieren können, dass die tatsächliche Leistungsvarianz nur unpräzise reflektiert wird (Vidulich, 2002) und außerdem die Messung von Scanning Techniken dadurch verzerrt werden kann. Da es in der Heeresfliegerwaffenschule üblich ist, Leistungsbewertungen subjektiv über die Komponenten der mentalen Restkapazität, der Abweichung vom optimalen Landepunkt und einer Beurteilung des Airmanship vorzunehmen, wurde dies auf die Eye Tracking Studie übertragen. Dabei müssen subjektive Einflüsse der Fluglehrer, die für die Leistungsbewertung zuständig sind, in Kauf genommen werden, weil in der Heeresfliegerwaffenschule keine objektiven Daten, wie beispielsweise Abweichungsquadrate von Flugrouten oder Landepunkten, vorliegen. Dass subjektive Expertenbeurteilungen aber hinreichend akkurat und konsistent sein können, wurde durch Studien aus dem Leh-

rer-Schüler-Kontext bereits nachgewiesen (Akkuratheit in Bezug auf tatsächliche Leistung bis zu $r = .85$, Coladarci, 1986; Konsistenz zwischen Experten bis zu $r = .97$, Borman, 1977). Darüber hinaus wird davon ausgegangen, dass Experten der Hubschrauberführung (Fluglehrer) über ein hinreichendes Maß an Metakognition über die eigene Leistung verfügen und damit eine relativ akkurate Selbsteinschätzung vornehmen können (vgl. Tsang, 2002; O'Hare, 2002 zu den Vorteilen von Expertise; unter Berücksichtigung des Moderators Metakognition kann die Akkuratheit der selbstbewerteten Leistung bis zu $r = .68$ betragen, Ehrlinger & Dunning, 2003).

5. Daneben wird die **Beanspruchung (Workload)** während einer Mission als messbarer Indikator der Informationsverarbeitung von Hubschrauberpiloten, hier insbesondere als Maß für das Verhältnis von aufgewendeten und verfügbaren Ressourcen im Arbeitsgedächtnis, betrachtet. Den Erkenntnissen von z. B. Gopher und Donchin (1986), Hart und Staveland (1988), Wickens und Hollands (1999), Buld (2000), Flemisch (2001) oder Rubio et al. (2004) zufolge reflektiert die Varianz in der subjektiven Beanspruchung unterschiedliche Aufgabenanforderungen. Workload-Unterschiede dienen also als Indikator für den Einfluss der Aufgabenanforderungen auf die Informationsverarbeitung. Für die Messung wird der NASA-TLX gegenüber anderen subjektiven Verfahren präferiert, weil er gute messmethodische und anwendungsökonomische Testeigenschaften (Validität, Reliabilität) aufweist und die Workload mehrdimensional erfasst. So fließen in den NASA-TLX beispielsweise auch erregungsbezogene (Effort) und affektive (Frustration) Anteile ein. Die Nachteile des NASA-TLX (Erfordernis der Intro- und Retrospektion, Abfrage von nur globalen Aufmerksamkeitsressourcen usw.) werden bewusst in Kauf genommen, denn das Verfahren zeigt hohe Korrelationen (r bis zu $.80$) mit physiologischen Workload-Indikatoren (HR, HRV) und wird daher als konvergent-valides Messverfahren befunden.
6. Dem Modell von Wickens (2009), den Ausführungen von Tsang (2002, zum domänenspezifischen Wissen) und den umfangreichen Erkenntnissen von Dixon et al. (1990), Bellenkes et al. (1997), Kasarskis et al. (2001 zur peripheren Wahrnehmung von Piloten), Matessa und Remington (2005) sowie Sullivan et al. (2011) zufolge ist von einem Einfluss der **Flugerfahrung** von Piloten auf die Informationsverarbeitung, hier insbesondere auf die visuelle Aufmerksamkeit und damit auf Scanning Techniken von Hubschrauberpiloten, auszugehen.
7. Den Ausführungen von Wickens (2009), Denning et al. (2003 zur Komplexität) und den zahlreichen Erkenntnissen von Svensson et al. (1997), Colvin et al.

(2003), Thomas und Wickens (2004) sowie DiNocera et al. (2007) zufolge ist von einem Einfluss der **Aufgabenanforderungen** auf die Informationsverarbeitung, hier insbesondere auf die visuelle Aufmerksamkeit und damit auf Scanning Techniken von Hubschrauberpiloten, auszugehen.

8. Die Forschungsbefunde zum Einfluss der Flugerfahrung und Aufgabenanforderungen sind teilweise heterogen und lassen bisher Varianzanteile in Ergebnissen zu Scanning Techniken unerklärt. Auf der einen Seite steigt die Tendenz zu Blickkonzentrationen auf zentrale Displays (Visual Tunneling), auf der anderen Seite wird von einer Zunahme des zufälligen Umherblickens (Visual Scanning Randomness) berichtet (vgl. z. B. Colvin et al., 2003; Thomas & Wickens, 2004; DiNocera et al., 2007). Auszugehen ist daher von einem **wechselseitigen Einfluss von Erfahrung und Anforderungen** auf die Informationsverarbeitung und Scanning Techniken von Hubschrauberpiloten.
9. Das Modell des **Situationsbewusstseins** (Endsley, 1995) beschreibt die Informationsverarbeitung von Piloten synonym zum Modell von Wickens (2009), wenngleich es detaillierter ist (z. B. Einfluss von externen Faktoren wie Automation oder die Betonung eines geteilten Situationsbewusstseins bei Operationen im Team). Das Konstrukt des Situationsbewusstseins ist nicht hinreichend differenziert; unklar bleibt, ob SA Prozess- oder Outputvariable bei der Informationsverarbeitung ist und es fehlt z. B. eine detaillierte Abgrenzung zu Aufmerksamkeit oder Leistung. Außerdem ist der Zusammenhang mit Workload nicht geklärt. Die klassische Messung über die SAGAT ist mit Unterbrechungen der Hauptaufgabe verbunden (Intrusivität) und gilt als nicht ausreichend sensitiv (vgl. Hughes et al., 1990; Sarter & Woods, 1991; Seitz & Häcker, 2008, zitiert nach Herczeg & Stein, 2012). Entsprechend der theoretischen Redundanz und methodischen Defizite wird das Konstrukt SA aus der Datenerhebung dieser Studie ausgeschlossen.

Eine Zusammenfassung der Modellvorstellung für die vorliegende Eye Tracking Studie liefert Abbildung 25, die mit einer den Schlussfolgerungen entsprechenden Nummerierung ausgestattet ist. Über die Einflüsse der Faktoren „Aufgabenanforderungen“ und „Flugerfahrung“ (linke Seite) auf die zu messenden Indikatoren „Scanning Techniken“, „Performance“ und „Workload“ (rechte Seite) werden im Anschluss multivariate Hypothesen aufgestellt. Da der Einfluss affektiver und motivationaler Faktoren auf die hier fokussierten Variablen bisher nicht empirisch angezeigt wird, bleiben diese Einflüsse im Arbeitsmodell hintergründig.

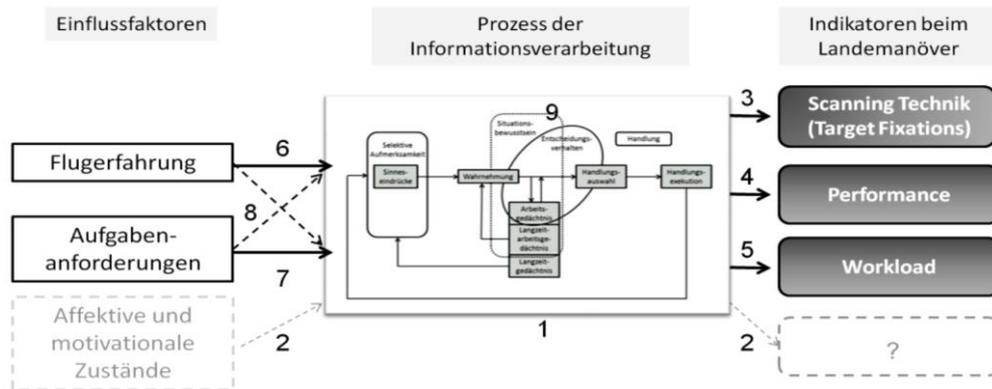


Abbildung 25: Arbeitsmodell für die Eye Tracking Studie.

2.3 Forschungsfragen und Hypothesen

2.3.1 Multivariates Modell

Geprüft werden voneinander unabhängige Nullhypothesen, die sich auf zwei Haupteffekte und eine Interaktion beziehen (in Anlehnung an Bortz, 1993, S. 287). Alle Hypothesen referieren auf die zu messenden Größen (abhängige Variablen):

a) Target Fixations, b) Performance und c) Workload,

die in den zu untersuchenden Gruppen jeweils verschiedene Gruppenerwartungswerte μ_{ij} erreichen. Die Gruppen (Untersuchungseinheiten) werden aus den beiden folgenden Faktoren (unabhängige Variablen) gebildet:

- a) Flugerfahrung mit den Faktorstufen hoch vs. gering ($I = 2$) und
- b) Anforderungen der Mission mit den Faktorstufen hoch vs. gering ($J = 2$)

Die Erkenntnisse zum **Einfluss der Flugerfahrung** auf Scanning Techniken von Luftfahrzeugführern zugrunde legend, wird davon ausgegangen, dass die Flugerfahrung von Hubschrauberpiloten sowohl Target Fixations, die Performance während des Fluges aber auch die Workload beeinflusst. Wenn die Flugerfahrung keinen Einfluss auf die abhängigen Variablen hat, dann gehören die unter den Stufen des Faktors Flugerfahrung beobachteten Untersuchungseinheiten Grundgesamtheiten mit gleichen Erwartungswerten an.

$$H_0^{\text{Flugerfahrung}} : \mu_{\text{unerfahrene Piloten}} = \mu_{\text{erfahrene Piloten}} \quad 35$$

Andernfalls läge ein Haupteffekt des Faktors Flugerfahrung vor.

$$H_1^{\text{Flugerfahrung}} : \neq H_0^{\text{Flugerfahrung}}$$

³⁵ Für die Notation gilt: Punkt (.) steht für „summiert über den jeweiligen Index aus der Faktorstufenkombinationsmatrix [Flugerfahrung x Anforderungen der Mission] i oder j“.

Basierend auf den dargestellten Erkenntnissen können Scanning Techniken von Hubschrauberpiloten, die Performance und Workload auch durch **Aufgabenanforderungen** determiniert werden. Wenn der Faktor Aufgabenanforderungen keinen Einfluss auf die abhängigen Variablen hat, gehören die Untersuchungseinheiten Grundgesamtheiten mit gleichen Erwartungswerten an:

$$H_0^{\text{Anforderungen der Mission}} : \mu_{\text{geringe Anforderungen}} = \mu_{\text{hohe Anforderungen}}.$$

Andernfalls läge ein Haupteffekt des Faktors Anforderungen der Mission vor.

$$H_1^{\text{Anforderungen der Mission}} : \neg H_0^{\text{Anforderungen der Mission}}.$$

Die Forschungsbefunde zum Einfluss der Flugerfahrung und Aufgabenanforderungen sind teilweise heterogen und lassen bisher Varianzanteile in Scanning Techniken, Performance und Workload unerklärt. Unklar ist, ob ein **wechselseitiger Einfluss von Erfahrung und Anforderungen** auf die abhängigen Variablen vorliegt. Wenn die beiden Faktoren die abhängigen Variablen nicht wechselseitig beeinflussen, dann setzen sich die Zellenerwartungswerte μ_{ij} additiv aus den Haupteffekten zusammen; kurz: Zwischen den Faktoren bestünde keine Interaktion.

$$H_0^{\text{Flugerfahrung x Anforderungen der Mission}} : \mu_{ij} = \mu_i + \mu_j - \mu.$$

Andernfalls läge ein Interaktionseffekt der Faktoren vor.

$$H_1^{\text{Flugerfahrung x Anforderungen der Mission}} : \neg H_0^{\text{Flugerfahrung x Anforderungen der Mission}}.$$

Üben die Faktoren Erfahrung und Anforderung einen signifikanten wechselseitigen Einfluss auf die abhängigen Variablen aus, wären signifikante Einflüsse der Einzelfaktoren obsolet (Janssen & Laatz, 2010, S. 377).

2.3.2 Explorative Fragestellungen

Aufgrund teils mangelnder Befundlage zu Scanning Techniken, zum Zusammenhang zwischen Leistung und Scanning Techniken sowie zur Nützlichkeit von Eye Tracking aus Sicht von Luftfahrzeugführern, sollen weitere Forschungsfragen explorativ untersucht werden, da sich hierzu keine Hypothesen ableiten lassen:

- Weichen objektive und subjektive Scanning Techniken in Abhängigkeit der Gruppenzugehörigkeit voneinander ab?
- In welche Richtung verschätzen sich Piloten hinsichtlich ihrer Scanning Techniken?
- In welchem Zusammenhang stehen die subjektive Leistung der Piloten und ihre Scanning Techniken?
- Wie hoch ist die Intrusivität und Nützlichkeit der Eye Tracking Methode aus Sicht der Piloten?

3 Methoden

Die folgende Darstellung der Methoden zur Hypothesenprüfung, die Dokumentation der Versuche sowie die Darstellung der Daten erfolgt anhand der ISO 15007, Teil 1 und 2. Hierin wird in Anhaltspunkten dargelegt, wie Versuchsequipment, Versuchsaufbau und Datenauswertung zu beschreiben sind.

3.1 Untersuchungsablauf und Experimentalplan

Bei der durchgeführten Untersuchung (s. Abbildung 26) handelt es sich um eine quasiexperimentelle Feldstudie, bei der das experimentelle Vorgehen in den realen Trainingsalltag der Piloten implementiert wird. Damit wird der Ausbildungsbetrieb auf der einen Seite relativ wenig interferiert, auf der anderen Seite sind dadurch Störeffekte, wie beispielsweise Absprachen zu Scanning Techniken unter den Probanden (Reaktivität) oder eine selektive Teilnahme von nur „bereitswilligen“ Probanden (Selektivität) nicht auszuschließen. In diesem Sinne ist die externe Validität des Versuches in Bezug auf die Grundgesamtheit der Hubschrauberpiloten hoch, wenngleich die interne Validität (Ceteris-Paribus-Validität) als eher niedrig einzustufen ist (vgl. Bortz, 1993, S. 747, 749).

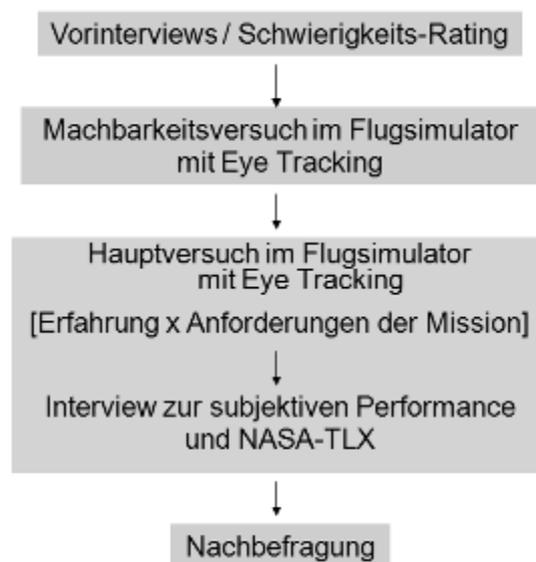


Abbildung 26: Untersuchungsablauf.

Der Hauptversuch war in einem 2 x 2 Faktordesign mit drei gemessenen Variablen geplant (s. Tabelle 3).

Tabelle 3: Experimentalplan.

Briefing & Kalibrierung des Systems				
Simulatorflug				
UV 1: Erfahrung	UV 2: Anforderungen	Gemessene Variablen	Gemessene Werte	Messverfahren
erfahrene Piloten	Mission 1 (geringere Anforderungen)	Target Fixations	...%	Eye Tracking
		Performance	Werte (subjektiv)	Interview
		Workload	...% (subjektiv)	NASA-TLX
	Mission 2 (höhere Anforderungen)	Target Fixations	...%	Eye Tracking
		Performance	Werte (subjektiv)	Interview
		Workload	...% (subjektiv)	NASA-TLX
	Nachbefragung		Werte (subjektiv)	Fragebogen
unerfahrene Piloten	Mission 1 (geringere Anforderungen)	Target Fixations	...%	Eye Tracking
		Performance	Werte (subjektiv)	Interview
		Workload	...% (subjektiv)	NASA-TLX
	Mission 2 (höhere Anforderungen)	Target Fixations	...%	Eye Tracking
		Performance	Werte (subjektiv)	Interview
		Workload	...% (subjektiv)	NASA-TLX
	Nachbefragung		Werte (subjektiv)	Fragebogen
Debriefing				

Anmerkung: UV = unabhängige Variable.

3.1.1 Machbarkeitsversuch

Vor dem Hauptversuch wurde eine Machbarkeitsstudie mit Fluglehrern der Heeresfliegerwaffenschule Bückeberg durchgeführt. Während der Erprobung waren die Piloten zunächst sowohl anhand halbstandardisierter Interviews (s. Anhang 7.2) als auch mithilfe eines Ranglistenverfahrens (s. Anhang 7.3) angehalten, über Ausbildungsabläufe im Simulator, während der Missionen stattfindende Scanning Techniken sowie die visuellen Anforderungen einzelner Missionen eine Expertenansicht zu geben. Auf den Aussagen des Experteninterviews aufbauend wurden zwei der Fluglehrer für einen Eye Tracking Vorversuch rekrutiert, während

dem sie zwei verschiedene Missionen flogen. Ziel des Vorgehens war es, die Implementierbarkeit des Eye Tracking Systems in den Simulator EC 135, die Vereinbarkeit des Systems mit der Fliegerausstattung der Hubschrauberpiloten (hier insbesondere Pilotenhelm, Verkabelung im Cockpit) sowie die Verwendbarkeit und Validität der Blickdaten zu überprüfen.

3.1.2 Hauptversuch

Vorbereitung. In einem zweiten Abschnitt wurde der Hauptversuch anhand des in Tabelle 3 dargestellten Experimentalplanes durchgeführt. Der Versuch fand in der Woche vom 16. bis 20.04.2012 im EC 135 Simulator Nr. 10 des Hans E. Drebing – Simulatorzentrums in Bückeburg statt. Die Befragung der Probanden mittels Interview und Fragebogen wurde beim Bundesministerium der Verteidigung (BMVg, P III 5) unter der Nummer 2/03/12 registriert. Im Vorfeld der Untersuchung wurden der Leiter des Simulatorzentrums, der zuständige Fliegerarzt sowie die potenziellen Probanden durch die Versuchsleiterin über zentrale Inhalte des Versuches ausführlich aufgeklärt. So wurden den Teilnehmern folgende Punkte erläutert (s. schriftliches Erläuterungsschreiben für alle Probanden im Anhang 7.5 nach den Vorgaben zur Durchführung von Untersuchungen in der Bundeswehr, VMBI vom 10.10.1990, S. 403 – 404):

- Ziel, Hintergrund und Ablauf der Studie,
- Versuchsequipment (Eye Tracking System, Interviews und Fragebögen),
- Anonymität jedes Teilnehmers sowohl während des Versuches als auch bei der Auswertung und Speicherung der Daten,
- Freiwilligkeit der Teilnahme,
- Nichtentstehung von Nachteilen bei Nichtteilnahme oder Ausscheiden aus dem Versuch und
- Speicherung der Daten.

Im Anschluss an die Erläuterungen wurden offene Fragen der Teilnehmer geklärt. Ein Proband fragte z. B. nach dem Ursprung der Auswahl der beiden Missionen. Entsprechend wurde durch die Versuchsleiterin noch einmal ausführlich von den Vorinterviews und der Rangliste der Anforderungen berichtet. In einem nächsten Schritt wurden die freiwilligen Probanden für den Versuchsablauf eingeteilt und solche Piloten rekrutiert, die die externe Konsole während der Simulatorflüge steuern und überwachen sollten. Ihre Aufgabe bestand während des Versuches darin, die jeweilige Mission im Simulator zu konfigurieren bzw. zwischen den Missionen hin- und her zu schalten. Sie waren mit den Versuchsteilnehmern und der Versuchsleiterin im Simulator stets über Funk verbunden. Vor und nach jedem Versuchstag wurde das Eye Tracking System mit seinen Elektronikkomponenten

im Simulator durch die Versuchsleiterin bzw. einen Assistenten auf- und abgebaut sowie auf Funktionalität geprüft. Die Komponenten wurden im Simulator gut sichtbar fixiert um ein Verrutschen während der Simulation mit Bewegung zu vermeiden (s. Abbildung 27).

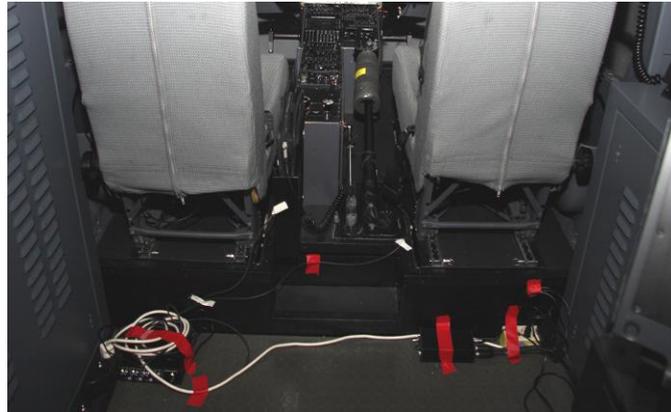


Abbildung 27: Fixierung der Komponenten des Eye Tracking Systems (eigene Fotografie).

Zu Beginn jedes Versuchstages kalibrierte die Versuchsleiterin das System zunächst mithilfe eines Assistenten, zu Beginn jedes Einzelversuches dann noch einmal mit jedem Proband (Details zur Kalibrierung s. Abschnitt 3.2.2).

Versuchsdurchführung. Nachdem jeder Proband mit der Eye Tracking Brille ausgestattet worden war, wurde eine der beiden Missionen geflogen. Sowohl der Faktor Erfahrung als auch die Anforderungen wurden dabei randomisiert gepaart, um Reihenfolgeeffekte zu vermeiden. Das heißt, dass immer abwechselnd ein Fluglehrer und ein Flugschüler an der Reihe waren und jeweils abwechselnd mit Mission 1 oder Mission 2 starteten (Beispiel: Proband 5 ist Fluglehrer und fliegt Mission 2 zuerst → Proband 6 ist Flugschüler und fliegt Mission 1 zuerst → danach fliegt Proband 5 Mission 1 → danach fliegt Proband 6 Mission 2 usw.). Zwischen jedem Flug hatte jeder Proband ca. 20 Minuten Pause während der nachfolgende Teilnehmer an der Reihe war. Dies geschah vor dem Hintergrund, Transfereffekte und Reaktivität unter den Teilnehmern zu vermeiden. Die Konventionen der Hubschrauberführergrundausbildung im Simulator, wonach Fluglehrer häufig links, Flugschüler häufig rechts im Cockpit sitzen, wurde während der Versuchsreihen beibehalten. Aufgrund nahezu vollständiger Symmetrie der beiden Cockpitseiten konnten Konfundierungseffekte zwischen Scanning Techniken und der Anordnung von Bedienelementen hierbei ausgeschlossen werden.

Nach jeder Mission erfolgte ein kurzes Interview durch die Versuchsleiterin über Funk. Fluglehrer wurden dabei um eine Selbsteinschätzung ihrer Scanning Techni-

ken, ihrer Performance und der subjektiven Workload während der Mission gebeten (s. Interview im Anhang 7.6). Auch die Flugschüler waren zu einer Selbsteinschätzung von Scanning Techniken und subjektiver Workload angehalten, ihre Performance jedoch wurde durch einen Fluglehrer bewertet (s. Anhang 7.7). Das Vorgehen findet seine Begründung in den in Abschnitt 2.2.8 dargestellten Erkenntnissen zur Akkuratheit von Expertenbewertungen. Darüber hinaus ist es während der Simulatoreausbildung ohnehin üblich, dass die Schüler durch die Lehrer beurteilt werden. Um hier Befangenheit der Fluglehrer zu vermeiden, gaben diese die Einschätzung ohne Zuhören des Schülers ab. Nachdem jeder Proband auf diese Weise zwei Missionen geflogen hatte, wurde ein Abschlussfragebogen zur Bewertung des Eye Tracking Verfahrens, des Komforts der Brille und der Eignung von Eye Tracking als Unterstützung für das Training (s. Anhang 7.8) ausgehändigt und durch die Probanden ausgefüllt. Durch dieses Vorgehen entstanden bei allen Teilnehmern Datensätze bestehend aus zwei Blickvideos für beide Missionen, zwei Interviews für beide Missionen und einem Abschlussfragebogen zur Bewertung des Eye Tracking Verfahrens. Die Blickvideos mit jeweiligem Interview konnten jetzt in vier Gruppen eingeteilt werden: Fluglehrer + Mission 1, Fluglehrer + Mission 2, Flugschüler + Mission 1, Flugschüler + Mission 2. Hiermit wurde nun – nach Bearbeitung der Rohdaten – eine hypothesengeleitete MANOVA realisiert. Nach Abschluss des Hauptversuches und Ausschöpfung der freiwilligen Teilnehmer führte die Versuchsleiterin ein Debriefing der Probanden durch: Es wurden Eindrücke des Versuches und das Vorgehen bei der Datenauswertung erörtert, die Probanden sahen ein Beispiel-Blickvideo und es wurden offene Fragen zur Berichterstattung geklärt.

3.2 Equipment

3.2.1 Simulator EC 135

Alle Untersuchungen wurden in Bezug auf das Hubschraubermuster EC 135 durchgeführt (s. Abbildung 28; links). Dies war der Tatsache geschuldet, dass Versuchsinhalte, wie zum Beispiel Abbildungen des Cockpits oder Ausbildungsinhalte publizierbar sind, während Cockpitaufnahmen anderer fliegender Waffensysteme der Bundeswehr der Geheimhaltung unterliegen (z. B. Eurofighter). Im Nachfolgenden werden die wichtigsten Daten bezüglich des Hans E. Drebing – Simulatorzentrums, indem sich der EC 135 Simulator (s. Abbildung 28; rechts) befindet, dargestellt (zur Verfügung gestellt von OTL Mallwitz).

Insgesamt betreibt die Firma CAE 12 Simulatoren (2 x UH-1D [VFR/IFR/BiV, CH-53 in Vorbereitung]; 2 x CH-53 [VFR/IFR/BiV], 8 x EC 135 [HGA]). Die Flugsimulatoren bestehen alle aus einem Basismodul, einem Bewegungssystem, einem Cockpitbe-

dienmodul, einer Dom-Konsole, einer externen Konsole, einer Debriefing-Station sowie einem Image Generator. Für jedes Hubschraubermuster stehen ein Cockpitmodul (Typenmodul), sowie eine typspezifische Software zur Verfügung.



Abbildung 28: EC 135 Hubschrauber³⁶ und Simulator-dom (eigene Fotografie).

Das Bewegungssystem der Simulatoren besitzt sechs Freiheitsgrade. Sie werden sowohl über ein Kniebrett im Flugsimulator als auch von außen, an einer Konsole gesteuert (s. Abbildung 29).



Abbildung 29: Externe Konsole (eigene Fotografie).

Mit den Flugsimulatoren können folgende Ausbildungsinhalte durchgeführt werden: Hubschrauberführergrundausbildung (inkl. BiV-Training sowie IFR), spezielle Musterausbildung, BiV- und IFR-Standardisierung. Weitere technische Daten zu den Flugsimulatoren sind Tabelle 4 zu entnehmen.

³⁶ http://de.wikipedia.org/wiki/Datei:GA_EC135_8257.JPG [Zugriff: 28.04.2013].

Tabelle 4: Technische Daten zum Flugsimulator EC 135.

Technische Parameter [Einheit]	Ausprägung im Simulator
Gewicht eines Domes [t]	13.5
Hydraulikdruck [bar]	115
Freiheitsgrade Bewegungssystem [df]	6
Anzahl Projektoren [n]	8
FOV [°]	240 x 90
Auflösung [Pixel]	1600 x 1200
Lichtstärke [Candela]	8
Anzahl Seatshaker [n]	2 (Pilot & Copilot)
Anzahl weiterer Sitzplätze [n]	2

3.2.2 Eye Tracking System

Hardware. Für die Datenaufzeichnung wurde das Head Mounted Eye Tracking System „Dikablis Cable“ der Firma Ergoneers verwendet. Die Head Unit (Kamerabrinne) des Systems wird in Abbildung 30 näher beschrieben.

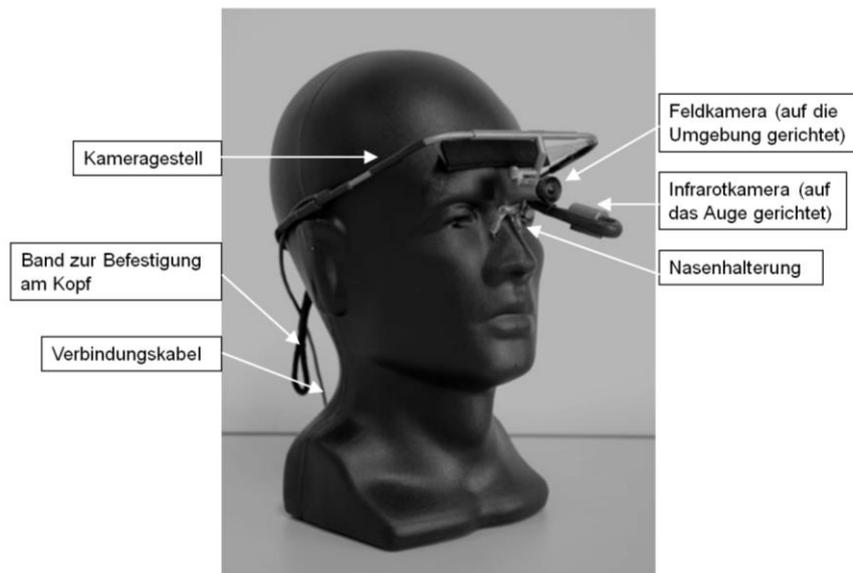


Abbildung 30: Head Unit³⁷.

Die Infrarotkamera erstellt ein Video des Auges (Eye Cam) und die Feldkamera (Field Cam) eines der Umgebung. Neben der Brille besteht das Eye Tracking Sys-

³⁷ <http://www.lfe.mw.tum.de/en/research/labs/gazetracking> [Zugriff: 28.04.2013].

tem Dikablis aus einer Elektronikeinheit (s. Abbildung 31, links), die die Videodaten der Kameras aufzeichnet und mit einem Rechner verbunden ist, welcher die Daten speichert. Zusätzlich zum Aufnahmerechner benötigt man außerdem Marker, die wahlweise aus Plastik oder als Infrarot-Marker für dunkle Umgebungen zur Verfügung stehen.



Abbildung 31: Elektronikeinheit und Infrarot-Marker (eigene Fotografien).

Die Marker ermöglichen der Auswertungs-Software, Blickdaten der Infrarotkamera mit Daten der Feldkamera abzugleichen und zu synchronisieren. Später können bei erkannten Markern AOIs eingezeichnet und Daten in deren Abhängigkeit ausgewertet werden. Bei der Messung muss mindestens ein Marker immer im Umgebungsvideo scharf zu erkennen sein. Folglich sind Vorversuche unter den jeweiligen Lichtbedingungen der Versuchsumgebung nötig. Für die vorliegenden Versuche wurde ein Infrarot-Marker im relativ wenig beleuchteten EC 135 Cockpit verwendet (s. Abbildung 31, rechts). Eine Zusammenfassung des Versuchsaufbaus mit den Hardware-Komponenten illustriert Abbildung 32.

Software und deren Einsatz während der Versuche. Tabelle 5 bietet einen Überblick über die benötigten Softwarekomponenten für einen erfolgreichen Eye Tracking Versuch mit dem System Dikablis Cable. Das System selbst enthält mehrere Software Features, die teilweise redundant sind (s. Handbuch der Firma Ergoneers GmbH). Insbesondere für die Kompatibilität der Daten und deren Weiterverarbeitung zu Kalkulationszwecken ist es nötig, weitere Programme hinzuzuziehen. Hier erwiesen sich Microsoft Excel (2003 für Windows) und SPSS (Version 17.0 für Windows) für Windows als hilfreich.

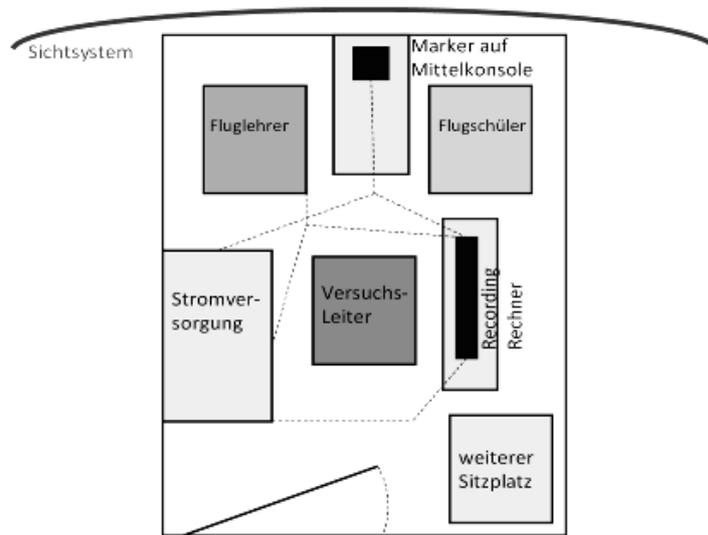


Abbildung 32: Skizzierter Versuchsaufbau (gestrichelte Linien: Kabelverbindungen).

Dikablis Recording ermöglichte vor jeder Videoaufzeichnung eine automatische Pupillenerkennung und Kalibrierung des Systems, welche jeweils manuell nachreguliert werden konnte (s. Abbildung 33). Die Pupille war immer dann durch das System erkannt, wenn sie als abgeschlossener Kreis rot (links unten) bzw. schwarz (rechts unten) abgebildet wurde. Mit Reglern (rechts oben) für die Einstellung der Pupillengröße, des Verhältnisses Pupille-Auge und zum Isolationsfaktor für Infrarot-Reflexion konnte die Erkennung verbessert werden.

Tabelle 5: Software-Komponenten.

Software	Funktion
Dikablis Recording V1.08	Kalibrierung und Datenaufzeichnung
Dikablis Player V1.08	Abspielen der Blickvideos
Dikablis Analysis V1.08	Nachbearbeitung und Optimierung der Blickvideos
D-Lab Marker Detection 2.1	Identifikation von Markern in Blickvideos
D-Lab Auswertungs-Software 2.0.1	Generierung und deskriptive Analyse der Blickdaten
Microsoft Excel 2003	Export der Blickdaten in Tabellenform
SPSS 17.0	inferenzstatistische Analyse der Blickdaten

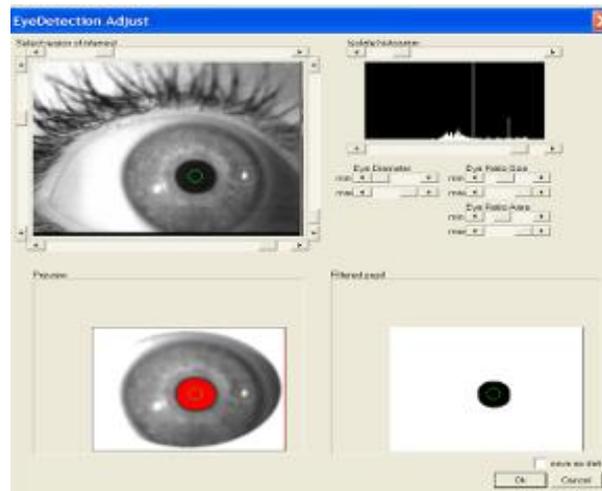


Abbildung 33: Aufnahmesoftware während Pupillenerkennung (aus: Dikablis Recording).

Bei der Kalibrierung handelt es sich um den horizontalen und vertikalen Abgleich bzw. die Synchronisierung zwischen Feld- und Infrarotkamera, sodass der tatsächliche Blickpunkt des Probanden im Blickvideo korrekt dargestellt wird. Hierzu wurden vor jedem Versuch im Blickfeld der Probanden drei feste Kalibrierpunkte vorgegeben (via weißem Klebeband), auf die ohne Kopfbewegung abwechselnd geblickt werden sollte (Punkte 1 – 2 = horizontale Kalibrierung; Punkte 2 – 3 vertikale Kalibrierung, s. Abbildung 34). Die Versuchsleiterin hat das System währenddessen mithilfe von Distanzreglern (Aufnahmesoftware) kalibriert. Außerdem ermöglichte es Dikablis Recording die Pupillenerkennung während der Datenaufzeichnung zu regulieren bzw. bei Kopfbewegungen zu verbessern. Nach abgeschlossener Kalibrierung konnte die Aufzeichnung gestartet werden.



Abbildung 34: Kalibrierpunkte (eigene Fotografie).

Ein fertiges Blickvideo wurde mithilfe der Software Dikablis Player abgespielt (s. Abbildung 35).



Abbildung 35: Dikablis Player.

Dikablis Analysis (s. Abbildung 36) wurde dazu verwendet, das aufgezeichnete Blickvideo zu bearbeiten und zu **optimieren**. Blickvideos konnten hiermit bei Bedarf geschnitten, die Pupillenerkennung im Nachhinein verbessert und das System offline noch einmal recalibriert werden.

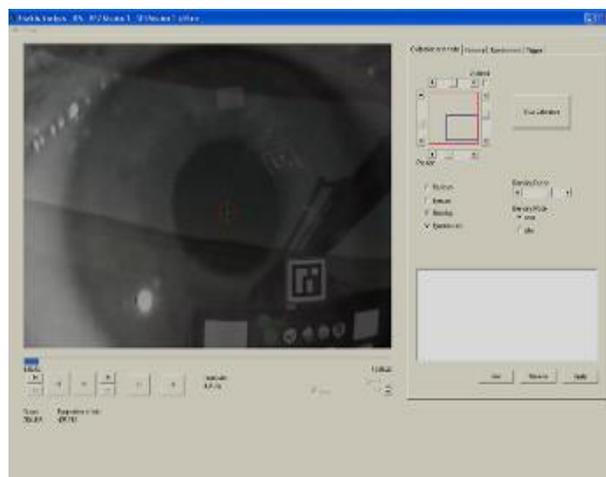


Abbildung 36: Dikablis Analysis während der Pupillenerkennung.

Die Software D-Lab Marker Detection erkennt die Marker in Blickvideos post hoc, indem die einzelnen Videosequenzen abgetastet und auf die Erkennbarkeit von Markern geprüft werden. Die Software meldet mit „Done“, dass die Markerdetek-

tion abgeschlossen ist. Erst nach der Identifikation der Marker wurden die Blickvideos in die D-Lab Auswertungs-Software importiert.

D-Lab (s. Abbildung 37) greift eng mit der Dikablis Software zusammen (s. Handbuch der Firma Ergoneers GmbH) und ermöglicht die grafische und tabellarische Darstellung der Daten. Im ersten Schritt wurde in D-Lab ein Validitätsindex für die Marker- und Pupillenerkennung berechnet. Lag der Index für die Pupillenerkennung bei $\geq 90\%$, wurden die Daten der Augenkamera als valide betrachtet. Die Güte der Markerererkennung wurde bereits bei einem Wert von $\geq 70\%$ (Feldkamera) als akzeptabel³⁸ interpretiert, weil eine noch höhere Erkennungsrate mit Infrarot-Markern bei schwacher Cockpitbeleuchtung aus praktischer Sicht nur schwer zu realisieren ist. Fälle, die die genannten Werte für beide Indizes unterschritten, wurden von der Datenauswertung ausgeschlossen. Im zweiten Schritt wurden AOIs als Flächen in die Blickvideos eingezeichnet, welche blau dargestellt werden. Für die durchgeführten Versuche wurden die AOIs Instrumente (für Head Down Blicke ins Cockpit) und Out The Window (OTW; für Head Up Blicke aus dem Cockpit heraus) gewählt. Trifft ein Blick in ein AOI, erscheint dieses in dunkelblau (s. Abbildung 37) in der Benutzeroberfläche. Im dritten Schritt wurde in D-Lab die Blickdauer in die AOIs berechnet (per Mausklick). Diese werden zunächst innerhalb der Benutzeroberfläche numerisch und grafisch dargestellt. Später werden sie vom System für weitere deskriptive Analysen verwendet. Im vierten Schritt wurden in den Blickdaten die Lidschläge (Blinks) per Mausklick eliminiert. Unter einem Blink werden Intervalle mit einer Spanne ≤ 120 ms zwischen zwei Blickzuwendungen (Handbuch der Firma Ergoneers GmbH) verstanden.

Danach wurden – ebenfalls per Mausklick – eine deskriptive Zusammenfassung der Daten, ein Blickverlaufdiagramm und eine Heat Map erstellt. Dies geschah zum Zwecke der visuellen Inspektion der Daten (eine Diskussion der Heat Map als potenzielles Auswertungstool findet sich in Abschnitt 2.1.4). Die Outputs sind allerdings in dem von D-Lab verwendeten Format nicht eins-zu-eins kompatibel mit anderen Softwares. Entsprechend mussten die Blick-Rohdaten mit Excel exportiert werden.

³⁸ Der Schwellenwert von 70 % wurde in Anlehnung an Nunnally (1978) gewählt. Der Autor schlägt als kritischen Wert für die interne Konsistenz von Items (Datenreliabilität) ein Cronbach's α von .70 vor, welches als „akzeptabel“ eingestuft wird. Da die interne Konsistenz ein Gütekriterium für die Zuverlässigkeit der Messung einer Variable mit mehreren Items ist, wurde der Wert von 70 % hier analog auf die Zuverlässigkeit der Erkennung von Markern über mehrere „Frames“ in einem Blickvideo übertragen.

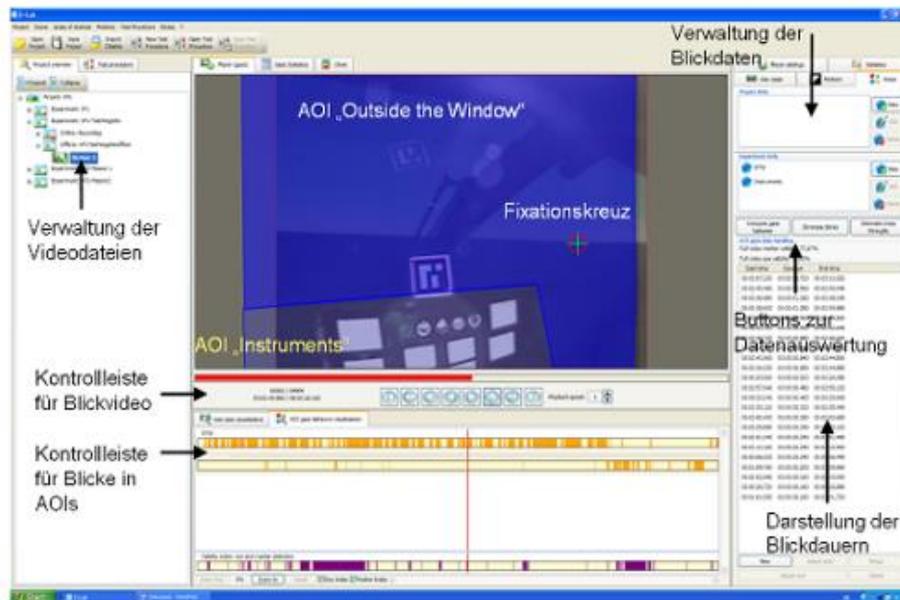


Abbildung 37: D-Lab Auswertungsoberfläche.

Beim Export der Blickdaten und inferenzstatistischer Analyse wurde aus D-Lab eine Tabelle mit der

- Dauer der Blickvideos,
- mittleren Blickdauer,
- maximalen und minimalen Blickdauer,
- kumulierten Blickdauer sowie
- Anzahl der Blicke

für jedes AOI übertragen. Bezogen auf Blickrohdaten wurde ferner eine Tabelle exportiert, die die exakte Dauer jedes einzelnen Blickes enthielt. Das Eye Tracking System tastet die Blicke der Probanden mit einer Frequenz von 40 Hertz ab. Nach Weitertransfer der Daten von Excel in SPSS konnten sie hier inferenzstatistisch ausgewertet werden (Details s. Abschnitt 3.4).

3.3 Operationalisierung der Konstrukte und zugehörige Untersuchungsinstrumente

3.3.1 Operationalisierung der unabhängigen Variablen

Flugführung. Die Erfahrung der Probanden wurde über ihre Flug- bzw. Simulatorstunden erfasst. Hierzu wurden sie wie folgt befragt (s. Anhang 7.8):

Flugstunden EC 135?	<input type="checkbox"/> 0 <input type="checkbox"/> 1-50 <input type="checkbox"/> 51-100 <input type="checkbox"/> 101-500 <input type="checkbox"/> 501-1000 <input type="checkbox"/> 1001-2000 <input type="checkbox"/> über 2000
Simulatorstunden EC 135?	<input type="checkbox"/> 0 <input type="checkbox"/> 1-50 <input type="checkbox"/> 51-100 <input type="checkbox"/> 101-500 <input type="checkbox"/> 501-1000 <input type="checkbox"/> 1001-2000 <input type="checkbox"/> über 2000

Als Gruppenvariable für die multivariate Datenauswertung wurden die Piloten bezüglich ihrer Erfahrung über die Funktion als Fluglehrer oder Flugschüler in der Heeresfliegerwaffenschule klassifiziert.

Anforderungen der Missionen. Um eine geeignete Operationalisierung für die Missionsanforderungen zu extrahieren, wurden Fluglehrer als Fachleute befragt. Im Rahmen eines halbstandardisierten Experteninterviews (s. Anhang 7.2) mit N = 6 Fluglehrern am 17.10.2011 wurden zusammenfassend die folgenden Aspekte geschildert. Die Fluglehrer gaben an, dass innerhalb der Hubschrauberführergrundausbildung (HGA) keine standardisierten Scanning Techniken vorgeschrieben oder definiert sind und jeder Pilot seinen eigenen Blickstil im Laufe der Zeit entwickelt. Grundsätzlich soll die Flugführung anhand der Außenwelt stattfinden. Ein zentrales Problem von Flugschülern sei es, dass diese zu unsystematischen Target Fixations neigten, indem sie Cockpitinstrumente oder Landmarks der Außenwelt zu lange fixierten und dabei zentrale Parameter der Flugführung (z. B. Flughöhe und Geschwindigkeit) vernachlässigten. Treten nach subjektiver Einschätzung der Fluglehrer falsche Scanning Techniken gehäuft bei einem Schüler auf, griffen diese zu „*drastischen Methoden, wie dem Abschalten der Displays im Simulator oder Real-Luftfahrzeug*“. Dies sei der Tatsache geschuldet, dass dysfunktionales Scanning den Trainings- und Ausbildungserfolg verlangsamt, die Flugsicherheit gefährdet und die Simulatorkrankheit begünstigen könne. Entsprechend sei es Ziel, den Schülern ein solches Blickverhalten so effizient wie möglich abzutrainieren. Beispielsweise bei den Manövern Autorotation oder Schwebeflug (Hovern) und Landen komme es zu vermehrten unsystematischen Target Fixations bei Schülern. Dies liege darin begründet, dass Missionen, die Schwebeflugelemente beinhalten, ein hohes fliegerisches und visuelles Geschick benötigen. Die Anforderungen von Missionen variierten also laut der Experten mit den visuellen Erfordernissen bei der Flugführung. Bei der Frage, was genau hohe Anforderungen einer Mission ausmache, antworteten die Fluglehrer wie folgt:

- Es ist schnelle Informationsaufnahme und -auswertung aus dem Cockpit erforderlich.
- Es muss eine große Datenmenge aus den Instrumenten bei gleichzeitig schwer einschätzbarer Umgebungsinformation (z. B. Festigkeit des Untergrundes, Abschätzen des Landewinkels) herangezogen werden.

- Es existieren kaum fixe Referenzpunkte im peripheren Sichtfeld.
- Es ist ein „Heranschweben“ an den Landepunkt erforderlich.
- Es sind Hindernisse bei begrenztem Landeplatz vorhanden.

Hinsichtlich der Frage, welche Landemanöver mit dem EC 135 Simulator grundsätzlich simuliert werden können, antworteten die Experten:

- Pinnacle (z. B. eine Kuppe beim Geländeflug, 5 x 5 Meter),
- Einsatzgruppenversorger (z. B. auf der Nordsee),
- Kammlinie (z. B. ein Damm beim Geländeflug),
- am Hang (z. B. beim Geländeflug) und
- eingegrenzte Fläche (z. B. ein Ausschnitt im Wald beim Geländeflug).

Im Anschluss an das Experteninterview wurde eine Rangliste der oben genannten fünf Landemanöver (s. Anhang 7.3) konstruiert und N = 15 Fluglehrern am 21. und 22.02.2012 zum Bearbeiten vorgelegt. Dies geschah vor dem Hintergrund, die durch halbstandardisierte Interviews in Erfahrung gebrachte Diskriminierung zwischen Landemanövern mit geringeren und höheren visuellen Anforderungen durch eine standardisierte empirische Befragung zu verifizieren. Die Fluglehrer waren angehalten, die fünf Manöver in eine Rangreihe von 1 = geringe visuelle Anforderungen bis 5 = höhere visuelle Anforderungen zu bringen. Es wurde statt einem Rating (gleiche Bewertungen können beliebig oft vergeben werden) ein Ranking (jede Bewertung kann nur ein einziges Mal vergeben werden) bevorzugt, weil „[...] diejenigen Befragten, die im Rahmen des Rating-Formats schwierigen Entscheidungen durch Nicht-Differenzierung ausweichen, in der durch das Ranking-Format definierten Forced-choice-Situation gezwungen werden, reflektierte Urteile abzugeben“ (Klein & Arzheimer, 2012, ohne Seite). Da in die Rangliste jeder Wert nur einmal eingeht, die Rangreihen aller Probanden mithin die gleiche Summe ergeben und ein arithmetisches Mittel der Daten deswegen nicht aussagekräftig erscheint, wurde die prozentuale Verteilung der Rangplätze der Landemanöver betrachtet (s. Tabelle 6). Im Ergebnis zeigte sich, dass die Landung auf dem Pinnacle von den meisten Probanden auf dem Rangplatz 1 (geringere visuelle Anforderungen) gesetzt wurde (fett markiert). Auf den letzten Rangplatz 5 (höhere visuelle Anforderungen) platzierten die meisten Probanden die Landung auf dem Gruppenversorger (fett markiert). Anzumerken ist hierbei, dass drei Probanden (20 % der Stichprobe) dieses Landemanöver noch nie geflogen haben, demnach also keinen Rangplatz vergaben.

Tabelle 6: Rangliste der fünf Landemanöver (N = 15).

Landemanöver	1 (geringe visuelle Anforderungen)	2 (wenig visuelle Anforderungen)	3 (mittlere visuelle Anforderungen)	4 (höhere visuelle Anforderungen)	5 (hohe visuelle Anforderungen)
Pinnacle	53.3	13.3	13.3	6.7	13.3
Kammlinie	0.0	20.0	40.0	40.0	0.0
Gruppenversorger	8.3	0.0	8.3	33.3	50.0
Hang	0.0	26.7	20.0	20.0	33.3
eingegrenzte Fläche	33.3	40.0	20.0	6.7	0.0

Auf Basis der Rangliste wurden die beiden Landemanöver Pinnacle und Gruppenversorger bezogen auf ihre visuellen Erfordernisse als distinkt betrachtet und für die Operationalisierung der Anforderungen verwendet. Es wurde vermutet, dass sie durch ihre unterschiedlichen Anforderungen an den Piloten klar differenzierbare Scanning Techniken hervorbringen würden. Einen Eindruck der Landemanöver gibt Abbildung 38. Die erste Mission (links: Pinnacle) beinhaltete eine Landung auf einer Art Kuppe im Gelände, welche aus dem Waldboden herausragte. Bei der zweiten Mission sollten die Piloten auf einem Einsatzgruppenversorger auf der Nordsee landen, wobei dieser sich nur leicht auf dem Wasser bewegte. Beide Missionen wurden bei Tag unter Schönwetterbedingungen mit Visual Flight Rules (VFR) absolviert. Der Unterschied zwischen den Manövern bestand darin, dass sich bei der Landung auf dem Pinnacle viele Referenzobjekte im peripheren Sichtfeld des Piloten befanden (Bäume, Strommasten), an denen er sich orientieren konnte. Die Landung war somit als relativ leichtes Flugmanöver anhand der Außenwelt für einen erfahrenen Fluglehrer zu bewältigen. Die Luftfahrzeugführer sollten eine kurze Platzrunde fliegen und danach landen (Dauer ca. 4 – 5 min). Im Gegensatz dazu gestaltete sich die zweite Mission anspruchsvoller, weil auf der Nordsee keine Referenzobjekte im peripheren Sichtfeld vorhanden waren. Entsprechend musste sich der Pilot durch geschickte visuelle Orientierung an den Cockpitinstrumenten an den Landepunkt „*heran hovern*“ (das heißt im langsamen Schwebeflug anfliegen; Dauer ca. 4 – 5 min). Der Landepunkt war laut Aussage der Fluglehrer dann als annähernd optimal zu betrachten, wenn wie folgt gelandet wurde:

- a) im letzten Drittel und etwa mittig auf dem Pinnacle, bzw.
- b) auf dem Mittelpunkt eines vordefinierten, aufgezeichneten Kreises auf dem Deck des Gruppeversorgers (s. Abbildung 39).



Abbildung 38: Missionen (links: Pinnacle, rechts: Gruppenversorger, aus: Dikablis Player).

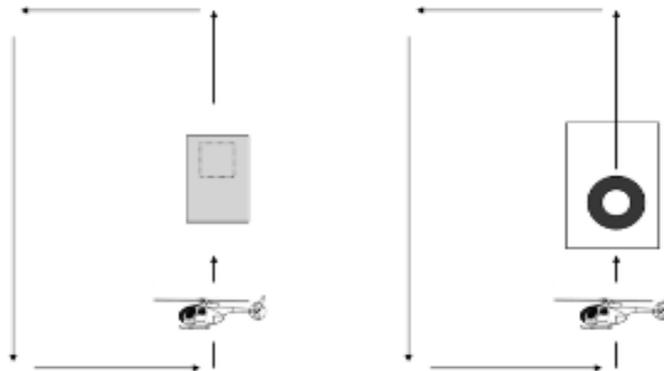


Abbildung 39: Platzrunde und Landepunkte (nicht maßstabgetreu³⁹).

Zusammenfassend wurden verschiedene Anforderungsgrade also über die visuellen Erfordernisse bei der Landung operationalisiert: Mission 1 (Pinnacle) wurde anhand der Expertenrangliste als Mission mit geringeren visuellen Anforderungen, Mission 2 (Gruppenversorger) als Mission mit höheren visuellen Anforderungen betrachtet.

3.3.2 Operationalisierung der abhängigen Variablen

Target Fixations. Die Variable Target Fixations wurde folgendermaßen operationalisiert: Ausgehend von den Empfehlungen der EASA (s. Abschnitt 2.2.6), wonach ein regulärer Blick ins Cockpit etwa 3.0 sec (= 3000 ms) dauern soll, wurden Blicke mit $X > 3000$ ms als Target Fixations ins Cockpit codiert. Demgegenüber wurde für einzelne Blicke in die Außenwelt (OTW) ein Berechnungsalgorithmus von Inhoff & Radach (1998) herangezogen, weil hierfür keine Empfehlung in Bezug auf die Hub-

³⁹ Grafikauszüge aus <http://www.rc-heli.cc/Assembler/media/images/OH-6A-Cayuse-schema.jpg> [Zugriff: 28.04.2013].

schrauberführung existiert und sich die Angaben der EASA für Airline-Piloten kaum auf Hubschrauberführer übertragen lassen. Hierbei galt: Wenn die Blickdauer X in einer Stichprobe mit einer Standardabweichung SD um einen Mittelwert M_x streuen, dann liegt die Blickdauer X mit $(M_x - 3*SD) \leq X \leq (M_x + 3*SD)$ im Normalbereich dieser Stichprobe, während die Blickdauer X mit $X > (M_x + 3*SD)$ Extremwerte abbildet, die um ein hohes Maß nach oben vom Mittelwert abweichen, Target Fixations implizieren und entsprechend codiert wurden. Der so aus den Blickdaten der Stichprobe ermittelte Wert wurde als Cutoff-Wert c für die OTW-Blicke der Probanden herangezogen (s. Tabelle 7).

Tabelle 7: Cutoff-Werte für Target Fixations.

	Cutoff-Wert c [ms]	Quelle
Instrumente	$c = 3000$	EASA (2010)
OTW	$c = M_{\text{Blickdauer}} + (3 * SD_{\text{Blickdauer}})$	Inhoff & Radach (1998)

Da ein zeitlicher Cutoff-Wert für Blickfixierungen in der Luftfahrt noch nicht empirisch abgesichert ist, wurden im Rahmen explorativer Folgeanalysen weitere Ausprägungen von c (2000, 4000, 5000 und 10000 ms) hinsichtlich ihres Effektes bei der multivariaten Varianzaufklärung getestet.

Performance. Alle Leistungseinschätzungen wurden auf Basis eines Experteninterviews subjektiv durch die Fluglehrer erteilt. Fluglehrer nahmen also eine Selbsteinschätzung vor (s. Abschnitt 2.2.8). Innerhalb von Interviews nach jedem Simulatorflug wurde die Leistung entsprechend der Angaben der Experten mit folgenden Items abgefragt (s. Anhang 7.6 und 7.7):

1. Wie viel Prozent betrug Ihre mentale Restkapazität während des Landemanövers? [%]
2. Wie hoch schätzen Sie Ihre Abweichung vom optimalen Landepunkt nach dem Landemanöver ein? [m]
3. Mit welcher Schulnote bewerten Sie Ihr „Airmanship“ (sicheres und vorausschauendes Führen des Luftfahrzeugs, Kommunikation) während des Landemanövers? [Schulnoten von 1 bis 5].

Die Variable Performance wurde bei der Datenauswertung wie folgt kompiliert:

$$\text{Performance} = (\text{Restkapazität} + \text{Airmanship}^{40}) - \text{Abweichung vom Landepunkt}$$

⁴⁰ Die Variable wurde umgepolt (rekodiert) und damit die Werte der Skala invertiert (Schulnote 1 → umgepolter Wert = 5, Schulnote 2 → umgepolter Wert = 4 usw.).

Nach dieser Gleichung ist die Leistung dann höher, wenn die Abweichung vom Landepunkt geringer ist (Restkapazität und Airmanship [umgepolt] konstant).

Workload. Zur Messung der subjektiven Workload nach jeder Simulatormission kam aufgrund seiner als gut befundenen diagnostischen Eigenschaften (s. Abschnitt 2.2.8) der NASA-TLX (Hart & Staveland, 1988) zum Einsatz. Für die vorliegende Untersuchung wurde der NASA-TLX leicht modifiziert (s. Anhang 7.6 und 7.7), da sich bei der bisherigen Anwendung zur subjektiven Beanspruchungsmessung bei Simulatorflügen (Stein & Robinski, 2011a, b, c) Auffälligkeiten des Verfahrens zeigten. Rein deskriptiv war beispielsweise offenkundig, dass über verschiedene Missionen und Stichproben hinweg die Einschätzung für Frustration sowie für mentale Anforderungen indifferent und der Gesamtwert stets – mit praktisch nur irrelevanten Abweichungen – um 50 % lag; dies, obwohl verschiedene Missionen verschiedene Grade an Gesamtbeanspruchung hervorbringen müssten. Hier könnten sowohl inhaltliche als auch methodische Ursachen zugrunde liegen. Rein inhaltlich ist die Vermutung gerechtfertigt, dass die Skalenbeschreibung zu Frustration falsch verstanden bzw. interpretiert wird; sie lautet zunächst übersetzt:

„Wie unsicher, entmutigt, irritiert, gestresst und verärgert (versus sicher, bestätigt, zufrieden, entspannt und zufrieden mit sich selbst) fühlten Sie sich während der Aufgabe?“

Für einen Fluglehrer beispielsweise könnte sich diese Fragestellung auf viele denkbare Aspekte beziehen, wie unter anderem Frustration deswegen, weil ein Schüler bestimmte Inhalte nicht umsetzt oder Frustration durch Grafikeigenschaften, die Simulator Sickness auslösen. Die eigentliche Frustration durch Handlungshemmnisse, welche die Gesamt-Performance beeinträchtigen und höhere Beanspruchung bedingen können, geht jedoch kaum aus dem Item hervor. Andere Skalen sind insbesondere für Simulationsflüge eventuell irrelevant, weil diese nicht in Aufgabenelemente zerlegbar sind. Am Beispiel der derzeitigen Fragestellung für zeitliche Anforderungen wird dies deutlich:

„Wie viel Zeitdruck empfanden Sie hinsichtlich der Häufigkeit oder dem Takt mit dem Aufgaben oder Aufgabenelemente auftraten? War die Abfolge langsam und geruhsam oder schnell und hektisch?“

Ein letztes Beispiel – die Skalenbeschreibung für Anstrengung – macht deutlich, wie wenig inhaltlich distinkt die Dimensionen des NASA TLX auf Probanden wirken könnten, sodass es ggf. zu Unschärfe in der subjektiven Bewertung kommt. Aus dem unten genannten Item geht nicht eindeutig hervor, ob es sich um mentale oder körperliche Anforderungen oder Anstrengung handelt:

„Wie hart mussten Sie arbeiten, um Ihren Grad an Aufgabenerfüllung zu erreichen?“

Aus methodischer Sicht ist denkbar, dass die Instruktion des NASA-TLX (sofern schriftlich vorhanden) unverständlich bzw. ungenau aufgrund fehlender Aufgabenspezifikation ist oder das vom Versuchsleiter vorgetragene Briefing bei insbesondere umfangreichen Versuchsplänen von Probanden vergessen wird. Wie aus obigen Beispielen hervorgeht, enthalten die Skalenbeschreibungen entgegen jeder Empfehlung zur Fragebogenkonstruktion (z. B. Dillman, 2000) zu viele Parallelaspekte (z. B. „unsicher, entmutigt, irritiert, gestresst und verärgert“), sodass Probanden nicht präzise antworten können. Für die vorliegende Befragung wurde der NASA-TLX zwar konventionell durchgeführt, um die Generalisierbarkeit der Daten und die Vergleichbarkeit der Ergebnisse mit bisherigen Studien zu gewährleisten, allerdings wurde zur Sicherstellung der Datenvalidität die Instruktion um eine aufgabenspezifischere Beschreibung der Subskalen leicht modifiziert. Die einzelnen Subskalen wurden während des Interviews mit folgenden Beispielen versehen (s. auch Anhang 7.6 und 7.7):

1. Mentale Anforderungen: z. B. Mehrfachaufgaben; Verarbeiten von viel Information aus dem Cockpit und der Umgebung,
2. Körperliche Anforderungen: z. B. Steuereingaben; Vermeidung von Simulator Sickness und Übelkeit; Ausgleichen der Bewegungsplattform,
3. Zeitliche Anforderungen: z. B. schnelle Abfolge von Manövern; Umsetzung von Missionsinhalten in zu kurzer Zeit,
4. Performance: z. B. geforderte Missionsinhalte erfüllt; Missionsinhalte gut umgesetzt,
5. Anstrengung: z. B. hohe oder lange Konzentration nötig, um Missionsinhalte umsetzen zu können,
6. Frustration: z. B. Grafik beeinträchtigt Training des Hovers oder Höheneinschätzung im Geländeflug; Simulator beeinträchtigt erfolgreiche Umsetzung von Missionsinhalten.

Innerhalb der Interviews nach den Missionen wurden den Probanden die Skalenbeschreibungen vorgelesen und ihre Antwort in Interviewprotokolle eingetragen. Bei der Datenauswertung wurde dann der Gesamtwert für Workload errechnet:

Workload = (mentale Anforderungen + physische Anforderungen + zeitliche Anforderungen + subjektive Leistung + Anstrengung + Frustration) / 6

Weitere Maße für die explorative Datenauswertung. Im Rahmen der halbstandardisierten Interviews zu oben dargestellten Variablen wurden die Probanden außerdem um eine **subjektive Einschätzung der eigenen Scanning Techniken**

gebeten (s. Anhang 7.6 und 7.7). Sie sollten beispielsweise einschätzen, wie viel Prozent ihrer Blicke sie nach außen bzw. auf die Instrumente richteten. Darüber hinaus waren sie angehalten, eine Einschätzung darüber zu geben, inwiefern eine sichere Flugführung durch ihre Scanning Techniken gewährleistet war. Diese Bewertungen dienten später als Indikatoren des Vergleiches objektiver und subjektiver Scanning Techniken. Zur Erfassung des subjektiven Komforts der Eye Tracking Brille während des Versuches und zur Bewertung der **Intrusivität** der Eye Tracking Prozedur (Einschränkungen der Sicht, Druckstellen durch Kameragestell oder Kabel, Störungen der Flugdurchführung durch Anwesenheit eines Versuchsleiters, Verstärkung von Simulatorkrankheit-Symptomen durch Eye Tracking) wurde den Probanden nach erfolgreicher Versuchsteilnahme ein Abschlussfragebogen (s. Anhang 7.8) zur Bearbeitung vorgelegt. Die **Einsetzbarkeit von Eye Tracking als Feedbacktool** in der HGA sowie im Realflug sollte durch die Probanden ebenfalls eingeschätzt werden. Für alle Items aus dem Fragebogen waren neben einem binären Format (ja/nein) auch offene Antwortmöglichkeiten vorhanden, um Kommentare und Ergänzungen der Probanden zu ermöglichen.

3.4 Datenauswertung

3.4.1 Hypothesengeleitete Datenauswertung

Zur hypothesengeleiteten Datenauswertung wurde eine **multivariate Varianzanalyse (MANOVA)** durchgeführt, deren Anwendung aus folgenden Gründen gerechtfertigt war: Um den Einfluss mehrerer (hier: zwei) unabhängiger Variablen auf mehrere (hier: drei) abhängige Variablen zu prüfen, bedürfte es ohne MANOVA mehrerer, univariater statistischer Tests (ANOVAs) mit derselben Stichprobe. Dies würde zu einer Kumulierung des Fehlers erster Art (α -Fehler) und damit zu einer höheren Irrtumswahrscheinlichkeit führen. Darüber hinaus können mit der MANOVA solche Gruppenunterschiede aufgedeckt werden, die sich aus Linearkombinationen der Variablen ergeben, denn das Verfahren berücksichtigt im Kontrast zur ANOVA auch Zusammenhänge zwischen den abhängigen Variablen. Aufgrund der Tatsache, dass die MANOVA auch Einflüsse durch Variablenkombinationen prüft, ist sie gegenüber der univariaten Varianzanalyse erschöpfender. Allerdings liefern die Testergebnisse im Falle von signifikanten Effekten keine klaren Einblicke, wo und wie genau die Gruppenunterschiede ausgeprägt sind. Dies macht Post Hoc Analysen notwendig (z. B. Scheffé-Test oder Follow Up ANOVAs).

Die MANOVA prüft Unterschiede in den Mittelwertvektoren der abhängigen Variablen, indem die Varianz der Daten zwischen den Gruppen und innerhalb der Gruppen analysiert wird. Die als hypothetischer Vergleich dienende Kovarianzmatrix der Grundgesamtheit wird über die Varianz der Mittelwertvektoren um

den Gesamtmittelwertvektor sowie über die Kovarianzmatrix innerhalb der Gruppen geschätzt, wobei hier die Korrelationen der abhängigen Variablen einfließen. Damit eine MANOVA durchgeführt werden kann, müssen folgende statistische Voraussetzungen erfüllt sein (Bortz, 1993, S. 317):

1. Multivariate Normalverteilung: Es wird angenommen, dass alle abhängigen Variablen und alle ihre Linearkombinationen univariat und alle Teilmengen der abhängigen Variablen multivariat normalverteilt sind. Hierfür gibt es allerdings keinen direkten statistischen Test. Entsprechend muss diese Voraussetzung über die Prüfung der univariaten Normalverteilung (Kolmogorov-Smirnov-Anpassungstest und visuelle Inspektion der Verteilungen) getestet werden. Die univariate Testung ist eine notwendige, aber nicht hinreichende Bedingung für die MANOVA.
2. Homogenität der Kovarianzmatrizen: Die Kovarianzmatrizen aller abhängigen Variablen müssen in allen Gruppen und für jede Kombination von Gruppenfaktoren homogen sein. Dies wird mit dem Box M-Test, welcher die Kovarianzmatrizen-Gleichheit prüft, getestet.

3.4.2 Explorative Datenauswertung

Zur Analyse der Fragestellungen, wie:

- Weichen objektive und subjektive Scanning Techniken in Abhängigkeit der Gruppenzugehörigkeit voneinander ab?
- In welche Richtung verschätzen sich Piloten hinsichtlich ihrer Scanning Techniken?
- In welchem Zusammenhang stehen die Performance der Piloten und ihre Scanning Techniken?
- Wie hoch ist die empfundene Intrusivität und Nützlichkeit der Eye Tracking Methode aus Sicht der Piloten?

wurden die Daten aus den Fragebögen anhand der objektiven Blickdaten gespiegelt. Es wurden Abweichungsvariablen gebildet und explorativ ausgewertet:

$$\text{Abweichung} = \text{subjektive Einschätzung [\%]} - \text{objektive Blickdaten [\%]}$$

Für alle explorativen Folgetests wurde die Irrtumswahrscheinlichkeit α nach Bonferroni adjustiert. Exploriert wurde darüber hinaus das Vorliegen linearer Zusammenhänge (Korrelations- und Regressionsanalysen) zwischen subjektiven und objektiven Daten.

3.5 Stichprobe

3.5.1 „Optimale“ Stichprobe

Die ex-ante Bestimmung des optimalen Stichprobenumfangs wurde mithilfe der Software GPower (Erdfelder, Faul, & Buchner, 1996) durchgeführt. Ausgehend von einem zweifaktoriellen, multivariaten varianzanalytischen Untersuchungsansatz, einer Mindestteststärke von Power $\geq .80$ (Cohen, 1969) und einer Irrtumswahrscheinlichkeit von $\alpha = .05$ wurden verschiedene Stichprobenumfänge für verschiedene Effektstärken f berechnet (s. Abbildung 40).

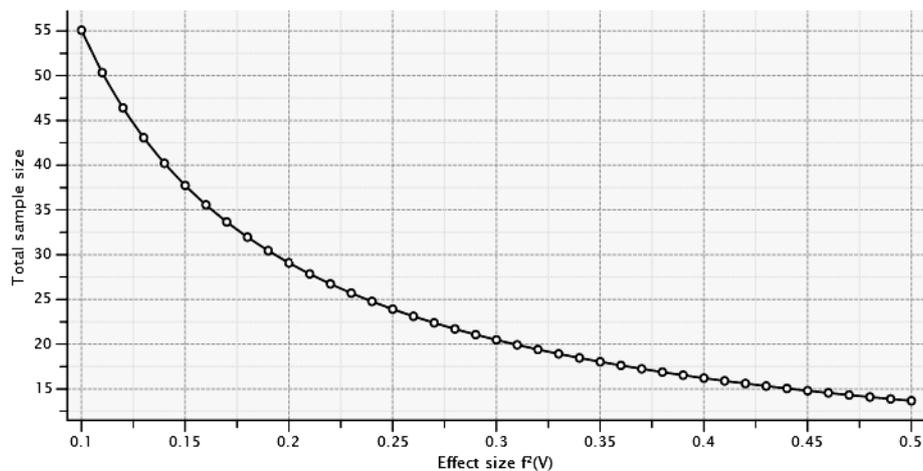


Abbildung 40: Ex-ante Bestimmung der Stichprobengröße.

Bei Vorliegen eines kleinen Effektes (Cohen, 1988) von $f = .10$ müssten unter den oben genannten Voraussetzungen 56 Probanden (Total Sample Size, s. Abbildung 40, N) untersucht werden. Dies entspräche bei vier Gruppen unter den Experimentalbedingungen einer Gruppengröße von $n = 16$. Zum Aufdecken eines mittleren Effektes von $f = .25$ bedürfte es $N = 24$ Probanden mit $n = 6$ pro Experimentalgruppe. Bei einer großen Effektstärke von $f = .40$ betrüge der optimale Stichprobenumfang $N = 20$ mit einer Gruppenstärke von $n = 5$. Im Vergleich dazu müssten unter Anwendung eines univariaten Vorgehens (ANOVA) zur Aufdeckung eines mittelgroßen Effektes $N = 128$ ($n = 32$) Probanden zur Verfügung stehen.

3.5.2 Tatsächliche Stichprobe

An der Untersuchung nahmen 33 männliche Piloten, die aus der Heeresfliegerwafenschule Bückeburg, III. Inspektion rekrutiert wurden, teil. Damit war die theoretische Voraussetzung zum Erreichen einer mittelgroßen bis großen Effektstärke erfüllt. Der zuständige Fliegerarzt vor Ort traf eine Vorauswahl der freiwilligen Teilnehmer nach der Maßgabe, etwa einen gleichen Anteil von Fluglehrern mit umfangreicher Erfahrung und Flugschülern mit weniger oder keiner Erfahrung zu

gewinnen. Von den 33 Probanden waren 16 Personen Flugschüler, die im Mittel etwa 51 – 100 Stunden Erfahrung sowohl im EC 135 Simulator als auch im Realfluffahrzeug besaßen. Die 17 Fluglehrer der Stichprobe hatten im Mittel 1001 – 2000 Stunden Flug- und 101 – 500 Stunden Simulatorerfahrung (Verteilung der Erfahrungskategorien s. Abbildung 41 und Abbildung 42).

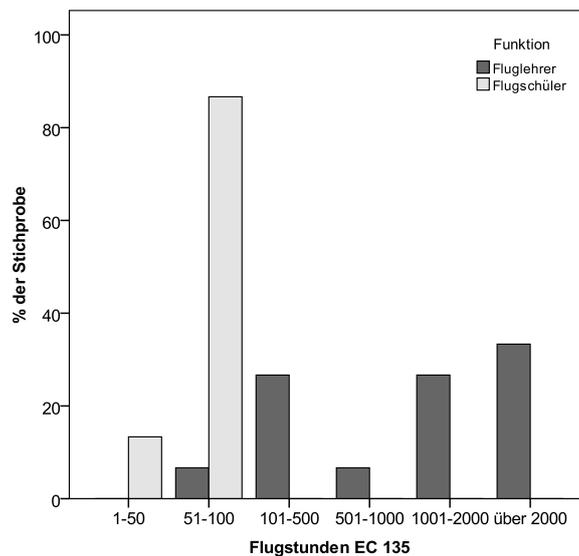


Abbildung 41: Flugerfahrung (Unterschied zwischen den Gruppen: $t[28] = 7.12, p < .000$).

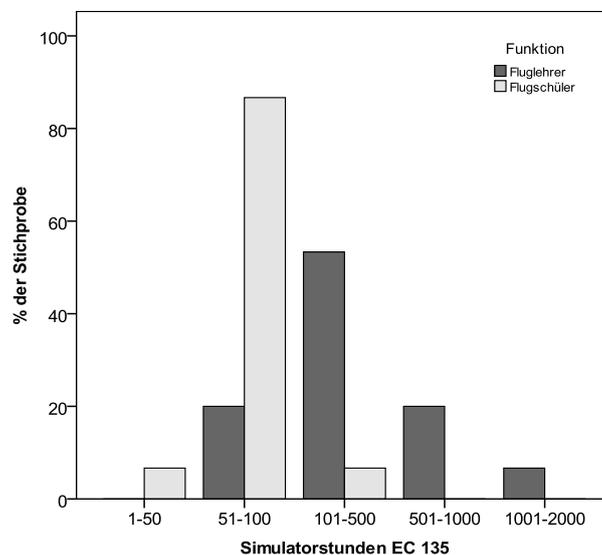


Abbildung 42: Simulatorerfahrung (Unterschied zwischen den Gruppen: $t[28] = 4.80, p < .000$).

Vier Teilnehmer trugen eine Brille, zwei trugen Kontaktlinsen; dies schränkte die Datenaufzeichnung nicht ein. Keiner der Teilnehmer hatte zum Zeitpunkt der Untersuchung akute Augenprobleme (z. B. tränende Augen, Bindehautentzündung). Entsprechend konnten die Blickdaten aller 33 Probanden erhoben werden.

4 Ergebnisse

4.1 Ergebnisse des Machbarkeitsversuchs

4.1.1 Implementierbarkeit des Eye Tracking Systems

Wie in Kapitel 3 dargestellt, wurde parallel zur Rangliste der Anforderungen mit $n = 2$ Fluglehrern (männlich, 50 und 52 Jahre alt) der Eye Tracking Machbarkeitsversuch durchgeführt. Hinsichtlich der Implementierbarkeit in den Simulator ergaben sich keine technischen bzw. Hardware-bezogenen Schwierigkeiten. Abbildung 43 bis Abbildung 44 geben einen Eindruck vom Versuchsaufbau.



Abbildung 43: Proband mit Brille (eigene Fotografien).

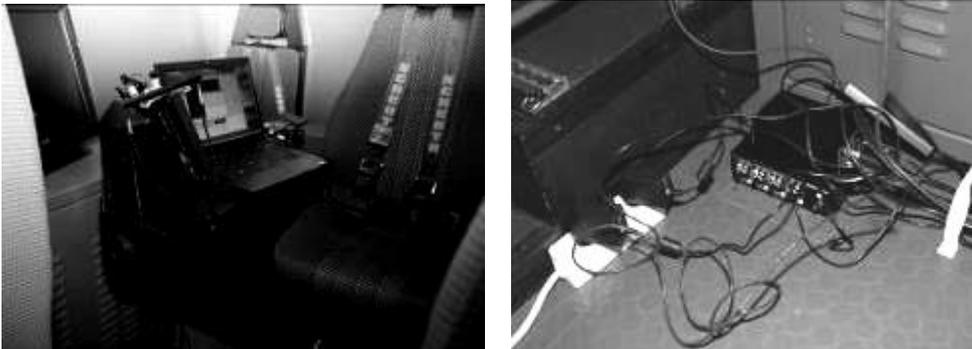


Abbildung 44: Versuchsleiterplatz und Verkabelung der Controlling Box (eigene Fotografien).

Als problematisch erwies sich allerdings nach jedem Versuchsdurchlauf das Tragen der Brille unter dem Pilotenhelm, was sich in Form von Druckstellen und Rötungen auf der Haut der Probanden zeigte. Außerdem verrutschte die Brille bei einem der Teilnehmer leicht (Abbildung 45), was eine Online- bzw. Nachjustierung der Pupillenerkennung und Kalibrierung erforderlich machte.



Abbildung 45: Verrutschen der Brille (eigene Fotografien).

4.1.2 Validität der Blickdaten

Tabelle 8 gibt einen Überblick über die Validität der Blickdaten (%) sowie über die mittlere Blickdauer (ms) und Blickanteile (%) auf Cockpitinstrumente und nach außen (OTW). Wie ersichtlich wird, waren die Validitätsindizes der Pupillen- und Markererkennung zufriedenstellend (> 90 % bzw. > 70 %).

Tabelle 8: Scanning Techniken (gemittelt) in AOIs.

VP	Mission	Dauer [min]	Marker-Validität [%]	Pupillen-Validität [%]	Ø Blickdauer Instrumente [ms]	Ø Blickdauer OTW [ms]	Ø Blicke Instrumente [%]	Ø Blicke OTW [%]
1	Pinnacle	11.78	79	98	524	2106	32.7	67.3
	Gruppenversorger	3.27	78	94	769	1884	29.2	70.8
2	Pinnacle	13.13	81	95	1655	2432	50.6	49.4
	Gruppenversorger	4.96	78	95	1339	2003	50.7	49.3

Anmerkung: VP = Versuchsperson, min = Minuten, Ø = Mittelwert, ms = Millisekunden (1 sec = 1000 ms), OTW = Out The Window.

Es erwies sich sowohl zwischen beiden Probanden eine Variabilität der Blickdaten als auch zwischen beiden Missionen: Die mittlere Blickdauer auf Cockpitinstrumente war erwartungsgemäß jeweils höher beim Landemanöver auf dem Gruppenversorger als auf dem Pinnacle. Ein reziprokes Muster zeigte sich für Blicke OTW. Bezogen auf die Variabilität zwischen den Piloten nutzte Proband 2 beide AOIs etwa gleichmäßig, während Proband 1 in beiden Missionen die Flugführung OTW bevorzugte.

4.1.3 Qualitative Auswertungsbeispiele für Einzelfallanalysen

Die Software D-Lab erlaubt eine grafische Auswertung des Nutzungsmusters der AOIs Instrumente (blau) und OTW (gelb). Wie aus Abbildung 46 hervorgeht, nutzte Proband 1 (oben) erst im letzten Teil der Landung auf dem Gruppenversorger die Instrumente verstärkt, während Proband 2 sowohl während der gesamten Platzrunde über der Nordsee als auch bei der Landung regelmäßig ins Cockpit schaute (unten). Bei der Abbildung ist zu beachten, dass grau dargestellte Abschnitte die Teile der Blickvideos repräsentieren, bei denen der Marker nicht erkannt und die Blickorte entsprechend nicht als solche ausgewertet wurden.

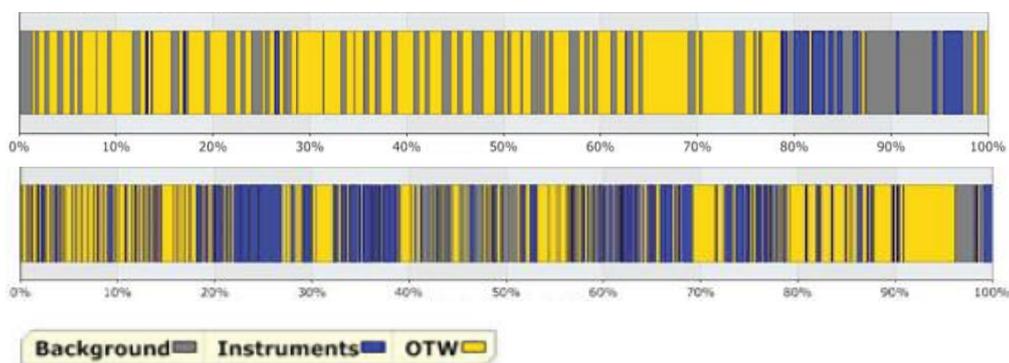


Abbildung 46: Proband 1 vs. 2 während Landung auf dem Gruppenversorger (aus: D-Lab).

Ferner war es mit D-Lab möglich, eine Heat Map zu extrahieren, mit der beide Missionen von einem Piloten hinsichtlich seiner Scanning Techniken vergleichbar waren. Anhand der Heat Maps für Mission 1 (Abbildung 47, links, geringere Anforderungen) und Mission 2 (Abbildung 47, rechts, höhere Anforderungen) ist die Relation zu erkennen, mit der Proband 1 die AOIs genutzt hat. Da die Heat Map weder linear skaliert noch numerisch interpretierbar ist, können damit keine quantitativen Aussagen gemacht werden. Zu entnehmen ist lediglich, dass der Proband in Mission 1 ein „breit gestreutes“ Blickmuster eher für OTW zeigte (rot), während er bei Mission 2 punktueller die Cockpitinstrumente nutzte (rot).

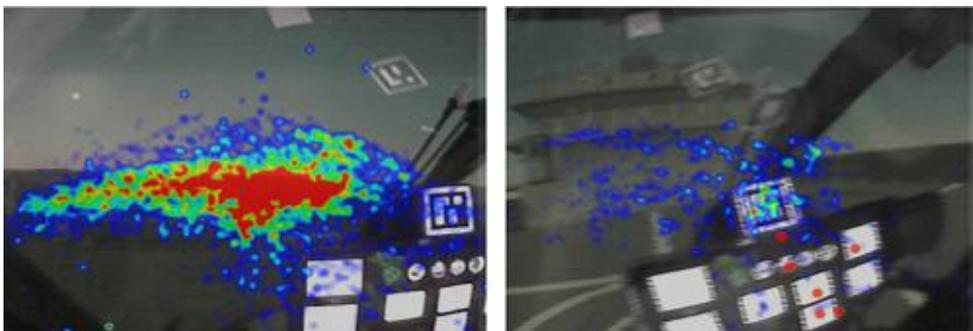


Abbildung 47: Heat Map für Proband 1 für Mission 1 vs. 2 (aus: D-Lab).

4.1.4 Zusammenfassung des Machbarkeitsversuchs

Wie Experteninterview und Rangliste erkennen lassen, war es für den Hauptversuch sinnvoll, die beiden Missionen Pinnacle und Gruppenversorger miteinander zu vergleichen, da sie laut Expertenmeinung jeweils eine Flugaufgabe mit geringeren bzw. höheren visuellen Anforderungen repräsentierten. Darüber hinaus erbrachte der Eye Tracking Machbarkeitsversuch, dass ein Eye Tracking System während der Simulatorflüge in das Cockpit implementierbar war. Allerdings deutete sich bereits im Machbarkeitsversuch an, dass die Brille zu Druckstellen unter dem Pilotenhelm führt und zum Verrutschen neigt. Entsprechend sollte das Halteband am Kopf der Probanden fester angezogen und auf eine gleichbleibende Position der Brille geachtet werden. Ein Problem bei Eye Tracking im Cockpit schien die Vereinbarkeit der Brille und des Visiers am Helm darzustellen, da Letzteres durch die Kamerabrille nicht geschlossen werden konnte. Beim Stand der Technik des Dikablis Systems kann dieses also vorerst nur bei (Simulator-) Flügen ohne Notwendigkeit des Visiers eingesetzt werden. Hinsichtlich der Markererkennung durch das System bestand insofern Verbesserungsbedarf, als dass die Validität der Detektion durch ein noch schärferes Einstellen der Feldkamera gesteigert werden sollte. Für eine quantitative Analyse der Scanning Techniken reichte die Software D-Lab allein nicht aus; hier mussten Excel oder SPSS herangezogen werden. Die qualitativen bzw. grafischen Analysemöglichkeiten aus D-Lab eigneten sich zwar nicht für inferenzstatistische Zwecke, könnten aber ggf. eine anschauliche Rückmeldung für Einzelfälle bieten. Wenn es gelingt, die Grafiken, wie Blickverlaufsdigramme oder Heat Maps, zeitnah zum Eye Tracking Versuch zu extrahieren, könnten diese als Debriefingmaterial ihre Nützlichkeit zeigen.

4.2 Ergebnisse des Hauptversuchs

4.2.1 Deskriptive Auswertung

Target Fixations. Von 33 Probanden lagen nach Durchführung des Hauptversuches 66 Blickdatensätze vor (33 x zwei Missionen), aus denen Target Fixations heraus partialisiert wurden. Vor der Extraktion der Target Fixations wurden die Blickdaten hinsichtlich ihrer Validität geprüft und mittlere Blickhäufigkeiten und -dauer innerhalb der Gruppen betrachtet (s. Tabelle 9).

Tabelle 9: Mittlere Blickhäufigkeiten und Blickdauer für OTW und Instrumente.

Erfahrung	Visuelle Anforderungen	n	Ø Mission-dauer [min]	Ø Blicke OTW [%]	Ø Blicke Instrumente [%]	Ø Blickdauer OTW [ms]	Ø Blickdauer Instrumente [ms]
hoch	gering	16	4.27	65.9	34.1	1520	880
	hoch	9	3.55	56.2	43.8	1140	880
gering	gering	10	4.74	51.7	48.3	1200	990
	hoch	10	4.18	54.4	45.6	1460	880
Gesamtstichprobe	--	45	4.21	58.2	41.8	1360	910

Anmerkung: n = Stichproben- bzw. Gruppengröße. Für die Gesamtstichprobe: Gewichtete Mittelwerte.

Es zeigte sich bei der Markererkennungsanalyse (s. Anhang 7.9), dass die Detektionsrate bei 21 von 66 Blickdatensätzen unter dem kritischen Wert von 70 % lag. Basierend auf dem Validitätskriterium wurden diese Blickdatensätze von der weiteren Analyse ausgeschlossen. Mithin konnten in die weitere Datenanalyse nur N = 45 Datensätze eingeschlossen werden. Im Gegensatz zur Validität der Marker war die Validität der Pupillenerkennung bei allen Probanden sehr hoch und lag bei über 90 % (s. Anhang 7.9). Wie aus Tabelle 9 hervorgeht, dauerte eine Mission im Mittel etwa vier Minuten, wobei das mittlere Verhältnis der Blicke nach außen (OTW) zu den Blicke auf die Instrumente bei etwa 60 : 40 % lag. Die Piloten setzten also im Gesamtdurchschnitt den fliegerischen Anspruch um, wonach die Flugführung vorrangig anhand der Außenwelt stattfinden soll. Bei Betrachtung der Blickdauer zeigte sich, dass Blicke nach außen im Mittel länger als eine Sekunde dauerten, Blicke auf die Instrumentierung hingegen unter einer Sekunde blieben.

Nach Bereinigung der Blickrohdaten und Inspektion der Verteilungen wurden aus den Blickdaten die Target Fixations extrahiert. Dies geschah, wie bereits ausführlich in Abschnitt 3.3.2 dargestellt, zum einen nach den Empfehlungen der EASA (2010) für Blicke auf die Instrumente, die maximal 3000 ms dauern sollen. Für Blicke nach außen wurde in der gesamten Stichprobe ein Cutoff-Wert nach Inhoff & Radach (1998) berechnet:

$$C_{OTW} = M_{Blickdauer} + (3 * SD_{Blickdauer})$$

$$C_{OTW} = 1360 \text{ ms} + (3 * 490 \text{ ms}) = \underline{2830 \text{ ms.}}$$

Die so errechneten 2830 ms wurden aus Gründen der Übersichtlichkeit auf 3000 ms gerundet und als Cutoff-Wert herangezogen. Im nächsten Schritt wurde

der Anteil der Blicke an der Gesamtzahl aller Blicke errechnet, der länger als 3000 ms dauerte. Außerdem wurde die prozentuale Verteilung dieser Target Fixations OTW bzw. auf die Instrumentierung betrachtet. Tabelle 10 illustriert, dass zwischen den Gruppen Unterschiede in der Auftretenshäufigkeit von Target Fixations existierten (M = 9.2 %, SD = 5.1 %). Im Durchschnitt dauerte eine Blickfixierung ca. 5700 ms bei einem empirischen Maximum von durchschnittlich 23000 ms. In der Gesamtstichprobe fanden Target Fixations eher bei Blicken nach außen im Vergleich zu Blicken auf die Instrumente statt (77 : 23 %).

Tabelle 10: Deskriptive Kennwerte für Target Fixations.

Erfahrung	Visuelle Anforderungen	n	Ø [%]	SD [%]	Ø Dauer [ms]	Range [ms]	Ø Anteil OTW [%]	Ø Anteil Instr. [%]
hoch	gering	16	10.7	5.1	5967	[3040-25080]	92.6	7.4
	hoch	9	6.3	4.3	5438	[3040-21560]	68.4	31.6
gering	gering	10	8.1	3.3	5702	[3040-24280]	68.0	32.0
	hoch	10	10.4	6.3	5637	[3040-21000]	80.4	19.6
Gesamtstichprobe	--	45	9.2	5.1	5686	[3040-22980]	77.4	22.6

Anmerkung: SD = Standardabweichung; Instr. = Instrumente.

Performance. Die Leistung nach jeder Mission wurde mit Performance = (Restkapazität + Airmanship [umgepolt]) – Abweichung vom Landepunkt errechnet. Die Spannweite des so gebildeten subjektiven Leistungsindikators reichte von [10.5 – 105.0]. Weitere statistische Kennwerte sind Tabelle 11 zu entnehmen. Wie in Tabelle 11 ersichtlich, war die mittlere Performanceeinschätzung erwartungsgemäß bei Fluglehrern höher als bei Flugschülern und innerhalb der Gruppen erzielten die Piloten bei der Mission mit geringeren Anforderungen (Pinnacle) die bessere Leistung.

Tabelle 11: Deskriptive Kennwerte für Performance.

Erfahrung	Visuelle Anforderungen	n	Ø	SD
hoch	gering	17	91.7	13.6
	hoch	17	81.7	20.2
gering	gering	16	72.4	20.6
	hoch	16	60.8	20.0
Gesamtstichprobe	--	66	77.0	21.7

Zur Inspektion der Parallelität der Performanceeinschätzung in der Heeresfliegerwaffenschule mit der Subskala „Performance“ des NASA-TLX wurden beide Variablen korreliert. Hierbei zeigte sich allerdings keine signifikante Korrelation. Gründe hierfür liegen entweder in einer zu geringen Streuung beider subjektiver Leistungsbewertungen ($M_{\text{Variable Performance}} = 76.96$, $SD = 21.67$; $M_{\text{NASA-TLX Performance}} = 83.26$, $SD = 15.45$, $N = 66$), darin, dass die Subskala „Performance“ des NASA-TLX nicht normalverteilt war ($KS-Z = 1.43$, $p = .035$) oder darin, dass das Performance-Item im NASA-TLX inhaltlich zu weit entfernt von der Leistungsbewertung in der Heeresfliegerwaffenschule ist. Während die Performance im NASA-TLX eher global mit „Wie erfolgreich haben Sie den Simulatorflug abgeschlossen?“ erhoben wurde, setzte sich die eigentliche Leistungseinschätzung aus mentaler Restkapazität, Abweichung vom Landepunkt und Airmanship summativ zusammen. Diesbezüglich zeigte sich, dass das Item zu „mentale Restkapazität“ und die Subskala „mentale Anforderungen“ des NASA-TLX signifikant negativ korrelierten ($r = -.36$, $p = .001$). Außerdem ergab sich eine signifikant positive Korrelation zwischen der Subskala „Performance“ des NASA-TLX und dem Item für „Airmanship“ ($r = .24$, $p = .025$). Da die Variable Performance mit dem Gesamtwert des NASA-TLX signifikant negativ korrelierte ($r = -.33$, $p = .003$) und subjektiv gute Performer damit erwartungsgemäß eine geringere subjektive Beanspruchung empfanden, ist von einer adäquaten konvergenten Validität der subjektiven Leistungseinschätzung in Bezug auf die hier verwendeten Anforderungen auszugehen.

Workload. Die oben genannten Unterschiede bezogen auf die Performance fanden sich nicht in analoger Weise für den Gesamtwert des NASA-TLX (s. Tabelle 12), welcher sowohl zwischen Fluglehrern und Flugschülern als auch zwischen beiden Missionen kaum variierte.

Tabelle 12: Deskriptive Kennwerte für Workload (NASA-TLX [%]).

Erfahrung	Visuelle Anforderungen	n	$\bar{\mu}$ [%]	SD [%]
hoch	gering	17	38.3	13.2
	hoch	17	41.3	11.1
gering	gering	16	42.4	12.5
	hoch	16	42.7	12.2
Gesamtstichprobe	--	66	41.2	12.1

4.2.2 Inferenzstatistische Auswertung

Prüfung der statistischen Voraussetzungen. Im Anschluss an die deskriptive Datenauswertung, bei der sich erste Hinweise auf Einflüsse durch Erfahrung und Anforderungen der Mission auf die Variablen Target Fixations, Performance und Workload andeuteten, wurden die Daten auf ihre Verwendbarkeit für eine MANOVA überprüft. Tabelle 13 illustriert die Ergebnisse der Tests zur Prüfung auf Normalverteilung und Kovarianzmatrizenhomogenität. Bei beiden Testverfahren gilt, dass die Voraussetzungen für die Durchführung einer MANOVA dann erfüllt sind, wenn die H_0 beibehalten wird, der Test also nicht signifikant wird. Aus der Tabelle geht hervor, dass dies für alle Variablen der Fall war, das heißt sie waren zumindest univariat normalverteilt (s. Abbildung 48) und die Kovarianzmatrizen homogen. Somit war die Durchführung der MANOVA indiziert.

Tabelle 13: Überprüfung der statistischen Voraussetzungen.

Test	Prüfgröße	Target Fixations	Performance	Workload
KS-Test (univariat)	KS-Z (p)	0.82 (.511)	1.09 (.183)	0.45 (.989)
Box M-Test (multivariat)	F [18, 4207] (p)		0.72 (.798)	

Anmerkung: KS-Test = Kolmogorov-Smirnov-Anpassungstest, KS-Z = Prüfgröße, p = Signifikanzniveau, Box M-Test = Boxscher M-Test, F = Prüfgröße.

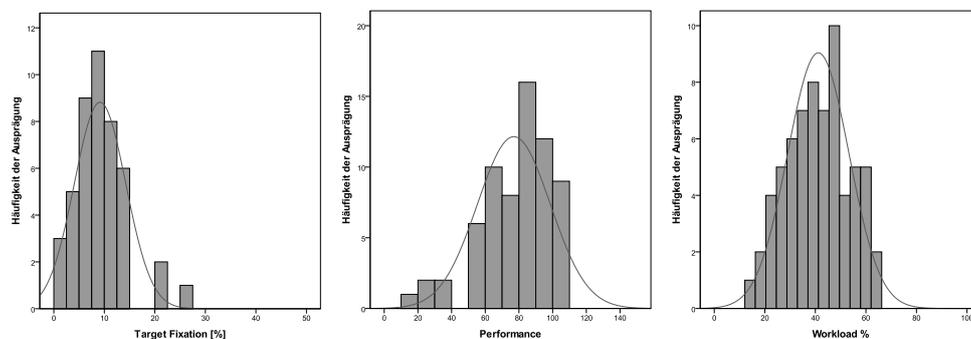


Abbildung 48: Univariate Normalverteilung der AVs (Target Fixations, Performance, Workload).

Ergebnisse der MANOVA. In Tabelle 14 sind die Ergebnisse der multivariaten Varianzanalyse, innerhalb derer der Einfluss der Faktoren geprüft wurde, dargestellt. Für den Faktor Erfahrung zeigte sich zwar erwartungsgemäß ein signifikanter Haupteffekt für die Variable Performance ($p < .000$) aber es ergaben sich entgegen den Erwartungen keine Haupteffekte für Workload oder Target Fixations. Der Faktor visuelle Anforderungen zeigte entgegen den Vermutungen ebenfalls keine signifikanten Haupteffekte für die abhängigen Variablen. In Kongruenz mit der

Annahme eines interaktiven Einflusses von Erfahrung und Anforderungen ergab sich aber ein signifikanter Interaktionseffekt für die Variable Target Fixations ($p = .033$), nicht jedoch für die Variablen Performance und Workload. Zur Post Hoc Analyse, wo genau die interaktiven Unterschiede für Target Fixations zwischen den Gruppen liegen, wurde die Abbildung 49 näher betrachtet. Außerdem finden sich im Anhang weitere quantitative Analysen (s. Anhang 7.11).

Während Flugschüler bei dem Manöver mit höheren visuellen Anforderungen mehr zu Target Fixations neigten, erwies sich für Fluglehrer ein umgekehrtes Muster: Sie zeigten bei der Mission mit geringeren visuellen Anforderungen mehr Target Fixations. Neben dem vermehrten Auftreten von Target Fixations bei Fluglehrern gegenüber den Flugschülern bei der Mission mit geringeren visuellen Anforderungen schätzten Erstere gemäß des oben genannten Haupteffektes ihre Gesamtleistung (Performance) in beiden Missionen signifikant besser ein (s. Abbildung 50).

Tabelle 14: MANOVA zur Hypothesenprüfung.

Quelle	Abhängige Variable	Quadratsumme	df	Mittel der Quadrate	F	p	Power
Erfahrung	Target Fixations	6.00	1	6.00	0.25	.623	.08
	Performance	5334.43	1	5334.43	15.64	.000	.97
	Workload	257.59	1	257.59	1.76	.193	.25
Visuelle Anforderungen	Target Fixations	12.45	1	12.45	0.51	.479	.11
	Performance	1229.43	1	1229.43	3.61	.065	.46
	Workload	148.15	1	148.15	1.01	.321	.17
Erfahrung x Visuelle Anforderungen	Target Fixations	119.23	1	119.23	4.88	.033	.58
	Performance	157.38	1	157.38	0.46	.501	.10
	Workload	8.42	1	8.42	0.06	.812	.06
Fehler	Target Fixations	1000.87	41	24.41			
	Performance	13980.79	41	341.00			
	Workload	6018.44	41	146.79			
Total	Target Fixations	4907.84	45				
	Performance	302972.00	45				
	Workload	82168.06	45				

Anmerkung: df = Degrees of Freedom.

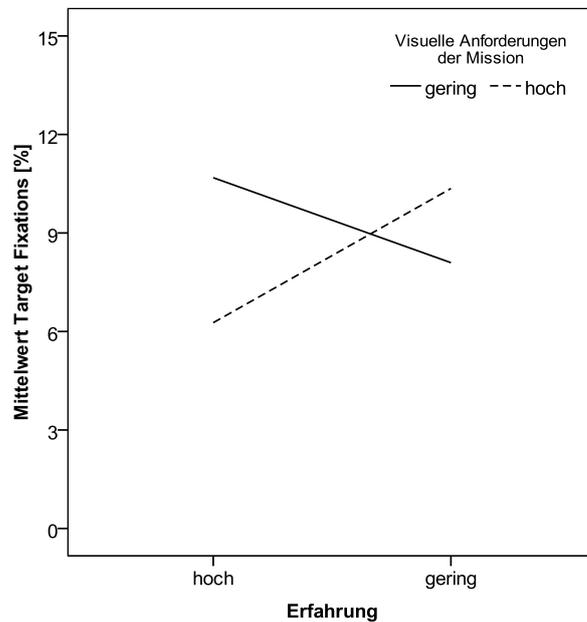


Abbildung 49: Interaktionseffekt der Faktoren für die Variable Target Fixations.

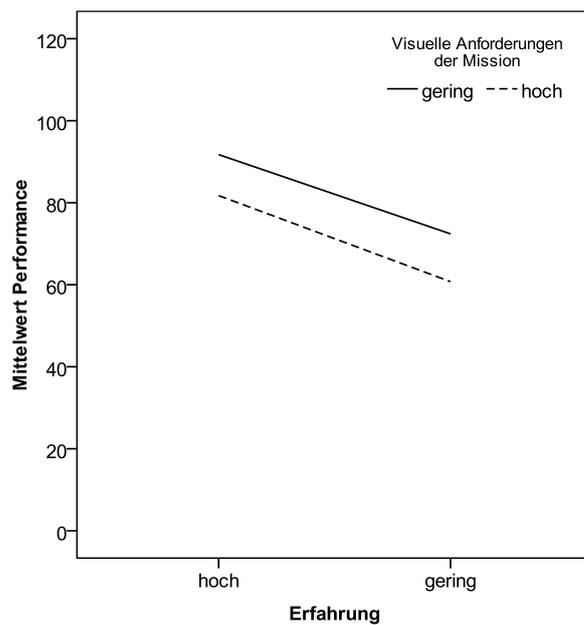


Abbildung 50: Haupteffekt des Faktors Erfahrung für die Variable Performance.

Bei einem Follow Up T-Test (α adjustiert) ergab sich auf dem Faktor Erfahrung außerdem eine signifikante Differenz für die Subskala mentale Anforderungen des NASA-TLX ($M_{\text{Lehrer}} = 45.6$, $SD = 23.8$; $M_{\text{Schüler}} = 60.6$, $SD = 24.1$; $T [64] = - 2.55$, $p = .013$; s. Anhang 7.10): Für Flugschüler war die subjektive mentale Beanspruchung bei beiden Missionen höher als für Fluglehrer. Dies galt aber offensichtlich nicht für die Gesamt-Workload.

Ergebnisse bei Variation des Cutoff-Wertes für Target Fixations. Da der zeitliche Cutoff-Wert für Target Fixations aus der hier vorliegenden Stichprobe errechnet wurde und es bisher keinen festgelegten Wert hierfür in der Luftfahrt gibt, wurde der Cutoff-Wert zusätzlich variiert (s. Tabelle 15). Mit der Verringerung (2000 ms) aber auch Erhöhung (4000 ms, 5000 ms, 10000 ms) des Schwellenwertes wurde geprüft, inwiefern sich die Aufklärungsgüte des multivariaten Modells verändert. Für alle anhand der Cutoff-Werte gebildeten Indikatoren waren die Voraussetzungen für die Durchführung einer MANOVA erfüllt (s. CD zur Dissertation, Flugmedizinisches Institut der Luftwaffe, Abteilung Ergonomie). Zur Schätzung des Beitrages zur Varianzaufklärung wurden die Konventionen von Cohen (1988) herangezogen, wonach $\eta^2 \geq .01$ einem kleinen, $\eta^2 \geq .06$ einem mittleren und $\eta^2 \geq .14$ einem großen Effekt entspricht.

Tabelle 15: Aufklärungsgüte der MANOVA bei Variation des Cutoff-Wertes.

Cutoff-Wert [ms]	η^2 gesamt	η^2 Target Fixations	Interpretation des Beitrags von Target Fixations (Cohen, 1988)
2000	.58	.15	Großer Effekt
3000	.54	.12	Mittlerer Effekt
4000	.51	.09	Mittlerer Effekt
5000	.49	.06	Mittlerer Effekt
10000	.48	.05	Kleiner Effekt

Anmerkung: η^2 = Effektstärke.

Bei der Interpretation der Modellgüte anhand von η^2 zeigte sich, dass bei Verringerung des Cutoff-Wertes eine größere Varianzaufklärung erzielt wurde als bei dessen Erhöhung. Ein kritischer Wert von 2000 ms entspräche dem sicherheitskritischen Maximum aus dem Fahrkontext mit Nebenaufgabenparadigma. Da sich die Blickdauer in der Luftfahrt nur schwer mit der aus dem Fahrkontext vergleichen lässt (nicht vergleichbare Geschwindigkeiten und Abstände von Referenzpunkten) und bei 3000 ms immer noch eine befriedigende Modellgüte erzielt wurde, erwies sich die Anwendung dieses Schwellenwertes als sinnvoll. Da beispielsweise ein Maximum von 10000 ms für eine maximale Blickdauer nicht mehr vergleichbar befriedigend (kleiner Effekt) zwischen den Gruppen trennt, ist seine Vorhersagekraft in den hier verglichenen Expertise- und Anforderungsgruppen offenbar geringer als die von 3000 ms.

Scanning Techniken in einzelnen Flugphasen. Bei der nachträglichen Analyse des gefundenen Interaktionseffekts wurde die Verteilung von Target Fixations anhand der Flugphasen Start – Platzrunde – Landung betrachtet. Hierzu wurde erneut eine MANOVA (s. Tabelle 16) durchgeführt, um den Einfluss von Erfahrung und

visuellen Anforderungen auf die Verteilung zu prüfen. Wie sich herausstellte, finden sich signifikante Unterschiede wieder auf der Faktorkombination für Target Fixations OTW in der Startphase ($p = .044$), während der Platzrunde auf die Instrumente ($p = .009$) sowie bei der Landung ($p = .030$). Zur Deskription wurden Abbildung 51 und Abbildung 52 herangezogen. Da die Verteilung von Target Fixations über Flugphasen betrachtet wurde, stand hier die verhältnismäßige gegenüber der absoluten Beschreibung im Vordergrund, das heißt die Abbildungen müssen auch untereinander vergleichend inspiziert und die Scanning Techniken der Flugschüler an denen der Fluglehrer gespiegelt werden. Die Interpretationsmöglichkeiten der gefundenen Scanning Techniken werden vertiefend in der Diskussion (Kapitel 5) durchgeführt.

Mission Pinnacle. Bei Start und Landung im Geländeflug (Pinnacle) führen Fluglehrer das Luftfahrzeug mehr anhand der Außenwelt als Flugschüler (dunkelblaue und dunkelgrüne Balken in Abbildung 51; Fluglehrer_{OTW} = 68.3 %, Flugschüler_{OTW} = 48.5 %) und fixieren während der Platzrunde weniger Informationen aus den Instrumenten (orange Balken in Abbildung 51; Fluglehrer_{Instrumente} = 3.1 %, Flugschüler_{Instrumente} = 22.3 %) als weniger erfahrene Piloten. Die Scanning Technik deutet auf einen Einsatz von Target Fixations der Fluglehrer in Flugphasen mit höherer Workload (Start, Landung) und gleichzeitig viel Umgebungsinformation hin. Flugschüler hingegen führen vor allem im Landeanflug offenbar eher kürzere Blicke durch. Dies wird anhand des geringeren Anteils an Target Fixations der Flugschüler insgesamt bei der Mission Pinnacle deutlich (11 % vs. 8 %, s. Tabelle 10).

Mission Gruppenversorger. Eine andere Scanning Technik der Experten liegt scheinbar bei der Landung auf dem Gruppenversorger vor. Während Flugschüler in der Startphase ihren Blick eher auf die Außenwelt fixieren, akquirieren Fluglehrer hier verhältnismäßig mehr Information aus den Instrumenten (blaue Balken in Abbildung 52; Fluglehrer_{OTW} = 7.9 %, Flugschüler_{OTW} = 18.8 %). Dies deutet darauf hin, dass Fluglehrer zunächst das Luftfahrzeug trotz der geringen Umgebungsinformation stabilisieren, indem sie in der Startphase eher Flugparameter aus den Instrumenten entnehmen bzw. überwachen. Während des Nordseeüberfluges dominieren bei den Experten – im Gegensatz zu den Schülern – dann Target Fixations in die Außenwelt (s. Verhältnis rote und orange Balken in Abbildung 52; Fluglehrer_{OTW} = 26.7 %, Flugschüler_{OTW} = 13.4 %), aber im Vergleich zum Geländeflug nutzen Experten auch hier vermehrt die Instrumente. Bei den Fluglehrern findet dann schon vor der Landung die Flugführung über der Nordsee offenbar mithilfe der vermehrten Fixierung des Gruppenversorgers (OTW) statt. Im Gegensatz dazu zeigen sich bei den Flugschülern erst bei der Landung mehr Target Fixations nach

außen, während sie beim Überflug der Nordsee eher kürzere Blicke in Außenwelt und Cockpit durchführen.

Tabelle 16: MANOVA für verschiedene Flugphasen.

Quelle	Abhängige Variable	Quadratsumme	df	Mittel der Quadrate	F	p	Power
Erfahrung	Start OTW	151.84	1	151.84	1.22	.276	.19
	Start Instrumente	3.97	1	3.97	0.02	.891	.05
	Platzrunde OTW	650.31	1	650.31	2.27	.140	.31
	Platzrunde Instrumente	651.98	1	651.98	3.49	.069	.45
	Landung OTW	162.23	1	162.23	0.48	.491	.10
	Landung Instrumente	13.06	1	13.06	1.39	.245	.21
Visuelle Anforderungen	Start OTW	6.27	1	6.27	0.05	.824	.06
	Start Instrumente	254.47	1	254.47	1.23	.274	.19
	Platzrunde OTW	1.82	1	1.82	0.01	.937	.05
	Platzrunde Instrumente	29.66	1	29.66	0.16	.692	.07
	Landung OTW	301.13	1	301.13	0.90	.349	.15
	Landung Instrumente	22.22	1	22.22	2.37	.131	.32
Erfahrung x visuelle Anforderungen	Start OTW	539.25	1	539.25	4.33	.044	.53
	Start Instrumente	36.97	1	36.97	0.18	.675	.07
	Platzrunde OTW	329.84	1	329.84	1.15	.290	.18
	Platzrunde Instrumente	1384.69	1	1384.69	7.41	.009	.76
	Landung OTW	1690.97	1	1690.97	5.04	.030	.59
	Landung Instrumente	1.98	1	1.98	0.13	.723	.06
Fehler	Start OTW	5106.19	41	124.54			
	Start Instrumente	8487.77	41	207.02			
	Platzrunde OTW	11771.14	41	287.10			
	Platzrunde Instrumente	7664.21	41	186.93			
	Landung OTW	13750.39	41	335.38			
	Landung Instrumente	384.33	41	9.37			
Total	Start OTW	14722.79	45				
	Start Instrumente	12874.06	45				
	Platzrunde OTW	30273.55	45				
	Platzrunde Instrumente	16683.76	45				
	Landung OTW	100512.86	45				
	Landung Instrumente	474.61	45				

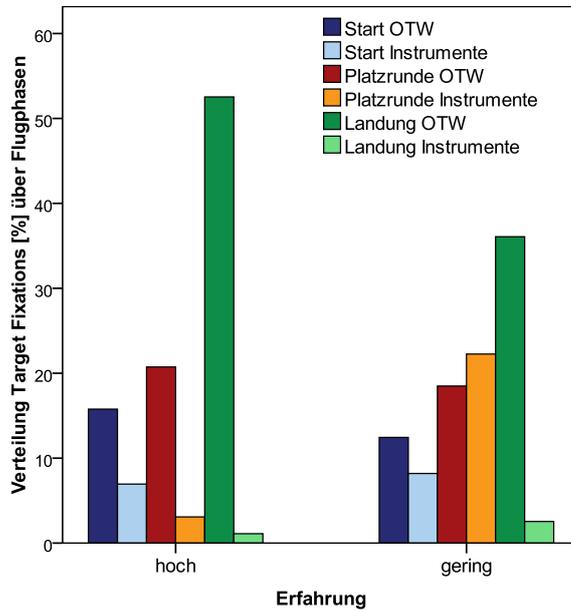


Abbildung 51: Relativer Anteil an allen Target Fixations über AOIs und Flugphasen für Pinnacle.

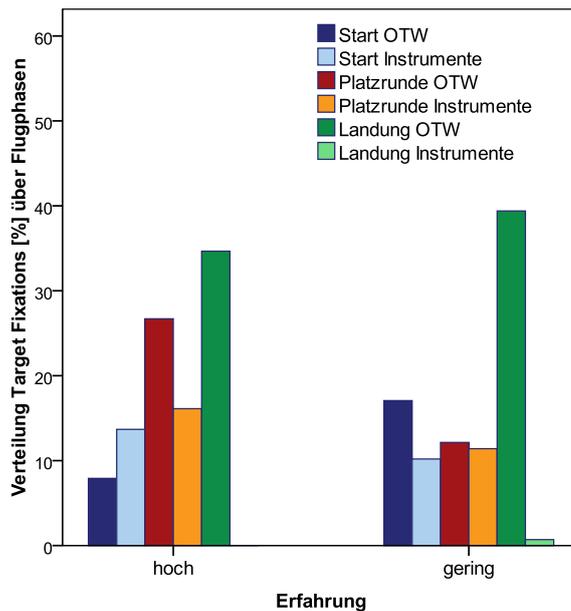


Abbildung 52: Relativer Anteil an allen Target Fixations über AOIs und Flugphasen für Gruppenversorger.

4.2.3 Explorative Auswertung

Abweichungen der subjektiven und objektiven Scanning Techniken. Zur Überprüfung der Abweichung subjektiver und objektiver Scanning Techniken wurden die aus $\text{Abweichung} = \text{subjektive Einschätzung [\%]} - \text{objektive Blickdaten [\%]}$ gebilde-

ten Differenzvariablen auf ihre Ausprägung und Verteilung inspiziert. Als Differenzvariablen entstanden somit:

1. Fehleinschätzung für Blicke OTW,
2. Fehleinschätzung für Blicke auf Instrumente.

In Bezug auf die oben dargestellte Formel zur Errechnung der Abweichung aus subjektiven und objektiven Scanning Techniken bedeuteten

1. negative Werte eine Unterschätzung der eigenen Scanning Techniken (z. B. es wurden durch Eye Tracking mehr Blicke gemessen als subjektiv geschätzt wurden) und
2. positive Werte eine Überschätzung der eigenen Scanning Techniken (z. B. es wurden durch Eye Tracking weniger Blicke gemessen als subjektiv geschätzt wurden).

Die Differenzvariablen erfüllten die statistischen Voraussetzungen für eine explorative MANOVA (s. CD zur Dissertation, Flugmedizinisches Institut der Luftwaffe, Abteilung Ergonomie) und waren univariat normalverteilt (s. Abbildung 53).

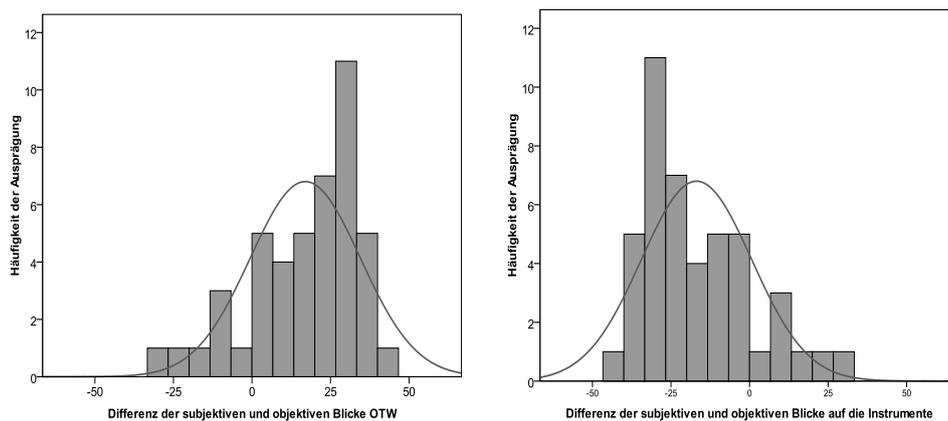


Abbildung 53: Differenzvariablen (links: OTW, rechts: Instrumente; N = 45).

Wie sich in den Grafiken zur Verteilung der Variablen bereits andeutet, lag für die Einschätzung der Blicke OTW im Mittel eine Verschätzung in positive Richtung vor, das heißt es wurden durch Eye Tracking weniger Blicke OTW gemessen als subjektiv geschätzt wurden (mittlere Differenz objektiver und subjektiver Daten = 17.0 %, SD = 17.6 %). Für Blicke auf die Instrumente lag ein entsprechend reziprokes Muster vor (mittlere Differenz = - 16.87, SD = 17.59).

Richtung der Fehleinschätzungen. Die explorative MANOVA auf den Faktoren Erfahrung und visuelle Anforderungen für die zwei oben genannten Abweichungsvariablen erbrachte einen signifikanten Haupteffekt auf Erfahrung für die Variablen Fehleinschätzung für Blicke OTW ($F = 10.1$, $p = .003$, $\eta^2 = .20$, Power = .87) und

Fehleinschätzung für Blicke auf Instrumente ($F = 10.6$, $p = .002$, $\eta^2 = .20$, Power = .89). Die Art und Weise, wie sich die Fehleinschätzungen über Fluglehrer und Flugschüler verteilten, wird in Abbildung 54 und Abbildung 55 dargestellt.

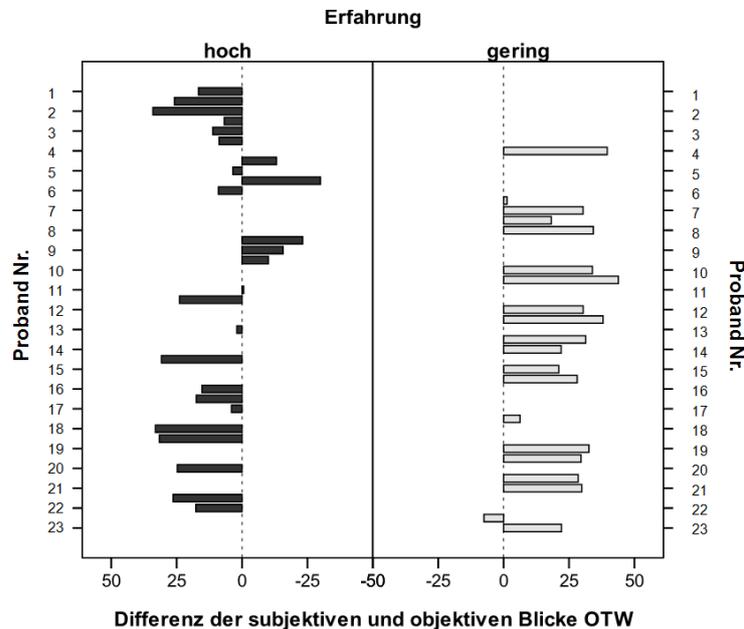


Abbildung 54: Fehleinschätzungen für OTW in Abhängigkeit der Erfahrung.

Die Abbildungen sind wie folgt zu interpretieren: Während die x-Achse die Differenzvariable mit dem Minimum von -50 % in der Mitte (!) repräsentiert, sind auf der y-Achse die einzelnen Probanden abgetragen. Im linken bzw. rechten Teil der Abbildungen befinden sich die beiden Erfahrungsgruppen. Je nachdem, in welche Richtung sich ein Proband hinsichtlich seines Blickortes verschätzt hat, ist ein „Ausschlag“ der Differenzvariablen zu verzeichnen. Wie die Grafiken illustrieren, verschätzten sich Flugschüler bezogen auf ihre OTW-Blicke fast ausschließlich in falsch-positive Richtung, das heißt sie überschätzten den tatsächlichen Anteil, mit dem sie nach außen geblickt haben um im Mittel 25.7 % (SD = 12.9 %). Fluglehrer verschätzen sich zwar ebenfalls eher in falsch-positive Richtung, dies aber nur mit im Mittel 10.0 % (SD = 17.9 %). Für die Einschätzung von Blicken ins Cockpit gilt wieder das umgekehrte Muster (Flugschüler: mittlere Differenz = - 25.8 %, SD = 12.9 %; Fluglehrer: mittlere Differenz = - 9.8 %, SD = 17.8 %). Das bedeutet, dass Blicke auf die Instrumente eher falsch-negativ eingeschätzt werden, also die Nutzungshäufigkeit unterschätzt wird (mehr Blicke gemessen als subjektiv geschätzt). Der geringere Abweichungswert von Fluglehrern ist unter anderem dadurch begründet, weil sich einige von ihnen in entgegengesetzter Richtung zu ihren Gruppenmitgliedern verschätzten (Summe positiver und negativer Werte nähert sich Null). Im Fazit lässt sich daher festhalten, dass sich auch erfahrene

Luftfahrzeugführer teilweise hinsichtlich ihrer Scanning Techniken verschätztten; hier die Varianz aber größer war als bei Flugschülern.

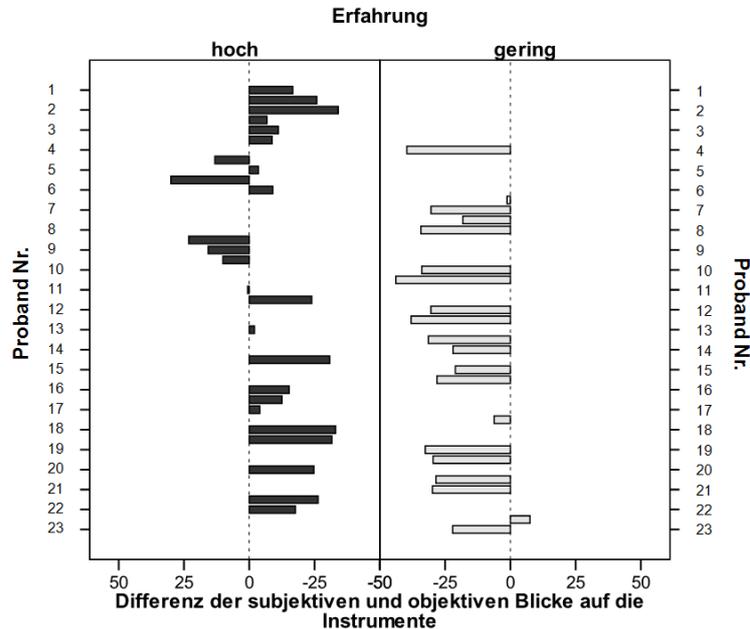


Abbildung 55: Fehleinschätzungen für Instrumente in Abhängigkeit der Erfahrung.

Zusammenhang von Scanning Techniken und Performance. Die explorative Korrelationsanalyse zur Detektion eines Zusammenhangs von Scanning Techniken und Performance wurde unter der Prämisse durchgeführt, Performance vorhersagen und einen möglichst großen Teil der Varianz erklären zu können. Es ergab sich allerdings kein signifikanter linearer Zusammenhang zwischen einzelnen Blickparametern (Target Fixations, Blicke OTW, Blicke auf die Instrumente) und Performance, wenngleich die Blickdaten untereinander korrelierten (s. Tabelle 17).

Tabelle 17: Korrelationen zwischen Performance und Blickparametern (N = 45).

	Performance	Target Fixations	Blicke Instrumente	Blicke OTW
Performance	1.00	.17 (.271)	-.19 (.205)	.19 (.205)
Target Fixations	.17 (.271)	1.00	-.41** (.006)	.41** (.006)
Blicke Instrumente	-.19 (.205)	-.41** (.006)	1.00	-.99** (.000)
Blicke OTW	.19 (.205)	.41** (.006)	-.99** (.000)	1.00

Anmerkung: (**) Die Korrelation ist auf dem Niveau von .01 (zweiseitig) signifikant.

Da Target Fixations ins Cockpit auf Basis der Empfehlungen der Luftfahrtbehörden abgeleitet wurden (> 3000 ms; EASA, 2010) und die Ergebnisse zeigen, dass die

subjektive Performance nicht mit dem Anteil der Target Fixations korreliert, lässt sich schlussfolgern, dass auch gute Performer von den Empfehlungen der Luftfahrtbehörden zum optimalen Scanning abweichen können. Damit sind die Empfehlungen der Luftfahrtbehörden offenbar keine notwendige Voraussetzung dafür, einen Hubschrauber sicher zu landen.

Die weitere Exploration erbrachte aber einen signifikanten linearen Zusammenhang zwischen Performance und der Fehleinschätzung der eigenen Scanning Technik: Bezogen auf den Instrument Check fand sich $r = .31$ ($p = .035$). Mit anderen Worten existiert offenbar ein Zusammenhang zwischen der Korrektheit der Selbsteinschätzung der Scanning Techniken und der Leistung beim Landemanöver.

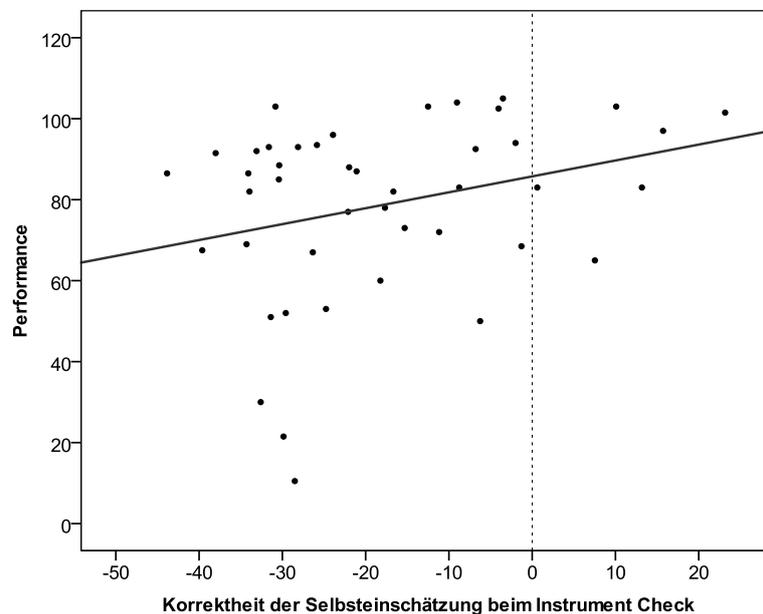


Abbildung 56: Regression Performance auf Selbsteinschätzung beim Instrumenten Check.

Soll die Performance aus der Selbsteinschätzung vorhergesagt werden, lässt sich folgende Regressionsgleichung anlegen: $Performance = 0.393 * X + 85.75$. Je näher die Abweichung zwischen objektiven und subjektiven Daten an Null liegt, desto größer ist offenbar die Performance. Da die Vorhersageleistung nur bei $R^2 = .10$ (s. Abbildung 56) liegt, stellt eine realistische Selbsteinschätzung von Scanning Techniken erwartungsgemäß nur einen Teil der Varianz der Performance dar.

Selbsteinschätzung von besseren und schlechteren Performern. Um zu verstehen, was bessere und schlechtere Performer bei den Landemanövern hinsichtlich ihrer Selbsteinschätzung unterscheidet, wurden solche Piloten herangezogen, deren Performance jeweils eine Standardabweichung über dem Mittelwert (Perfor-

mance ≥ 98.7) bzw. unter dem Mittelwert (Performance ≤ 55.3) lag. Als besonders gute Performer stellten sich $n = 9$ Fluglehrer heraus, zu den schlechteren ($n = 11$) zählten acht Flugschüler und drei Fluglehrer. Vergleicht man in Abbildung 57 die besseren Performer mit den schlechteren, zeigt sich, dass die Fehleinschätzung hinsichtlich der getätigten Blicke nach außen bzw. auf die Instrumente bei den besseren Performern signifikant geringer ausfällt (bessere Performer: $M_{OTW} = 4.5$, $SD = 17.6$; $M_{Instrumente} = -3.8$, $SD = 17.1$; schlechtere Performer: $M_{OTW} = 26.1$, $SD = 9.1$; $M_{Instrumente} = -26.1$, $SD = 9.2$; $F_{OTW} = 8.3$, $p = .014$; $F_{Instrumente} = 9.3$, $p = .010$). Diese Betrachtung untermauert die Unterschiede zwischen Fluglehrern und Flugschülern: Erfahrene und gute Performer schätzen sich hinsichtlich ihrer Scanning Techniken bewusster und realistischer ein.

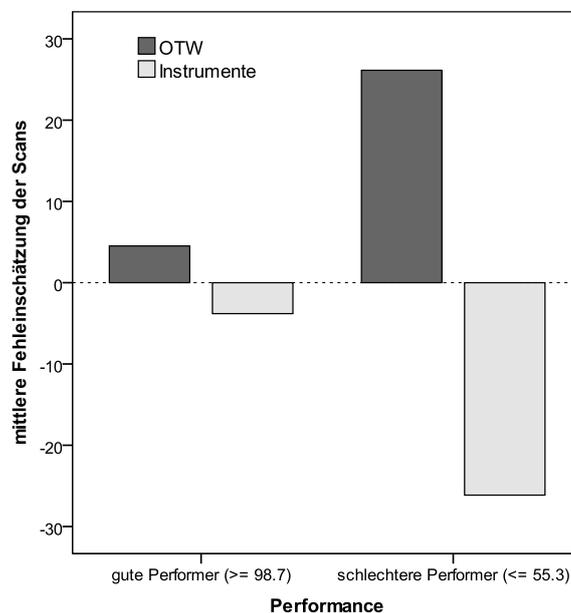


Abbildung 57: Selbsteinschätzung besserer ($n = 9$) und schlechterer Performer ($n = 11$).

Subjektive Intrusivität und Nützlichkeit der Eye Tracking Methode. Tabelle 18 zeigt die Ergebnisse zur Bewertung der Intrusivität und Nützlichkeit von Eye Tracking aus Sicht der Piloten. Dargestellt sind neben deskriptiven Prozentangaben der Gesamtstichprobe kritische Anmerkungen einzelner Probanden, die die offenen Antwortformate aus dem Abschlussfragebogen nutzten.

Hinsichtlich des wohl relevantesten Aspekts der sicheren Flugführung anhand der Außenwelt gab ein Drittel der Piloten eine eingeschränkte Sicht durch die Brille des Eye Tracking Systems an. Einzelfälle schilderten eine Beeinträchtigung des Sichtfelds für das Auge, vor dem die Augenkamera des Systems angebracht war, sowie eine Ablenkung durch das Kameragestell. Zwanzig Prozent der Stichprobe empfanden eine eingeschränkte Sicht auf die Cockpitinstrumente, welche für ein-

zelne Probanden mehr Kopfdrehungen als ohne Eye Tracking erforderlich machte. Dies galt offenbar besonders im Anflug bei der Landung und beim Cross Check der Instrumente. Die Tatsache der eingeschränkten Sicht durch das Kameragestell für einen Teil der Piloten ist sicherlich nicht vernachlässigbar bei der flugsicherheitsrelevanten Evaluation der Eye Tracking Methode. Allerdings relativierten sich diese Einschränkungen bei Betrachtung des Anteils von 83 %, der angab, sich an einer sicheren Flugführung durch Eye Tracking nicht gehindert zu fühlen.

Eher bedenklich waren die ergonomischen Eigenschaften der Brille, bezüglich derer 40 % der Stichprobe den Tragekomfort der Nasenauflage während des Fluges kritisierten. Immerhin ein Drittel der Piloten gab an, nach dem Flug unter Druckstellen auf dem Nasenrücken zu leiden (s. Abbildung 58).

Zusammenfassend beurteilten die Probanden den Komfort beim Tragen der Kamera-Brille unter dem Helm auf einer fünfstufigen Skala von 1 = optimal bis 5 = nicht akzeptabel mit $M = 3.07$ ($SD = 0.79$) und damit im Mittel als akzeptabel.

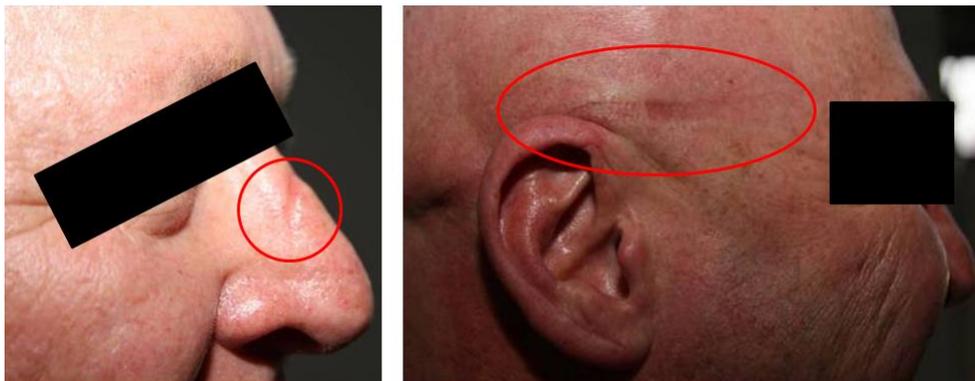


Abbildung 58: Druckstellen nach Tragen der Brille (eigene Fotografien).

Ein weiterer Fokus der Auswertung lag bei der Nützlichkeitsbewertung von Eye Tracking als Feedbackmethode beim Simulatortraining und seine Eignung für den Realflug. Über die Hälfte der Stichprobe konnte sich vorstellen, dass Eye Tracking regelmäßig einmal pro Monat als Unterstützung beim Simulatortraining eingesetzt wird (s. Abbildung 59). Einzelne Probanden wünschten sich aber eine Verbesserung des Gerätes und es wurde vorgeschlagen, ein System zu verwenden, welches außen am Helm fixierbar ist. Gegner der Methode wiesen beispielsweise auf den Zeitaufwand durch den Versuchsaufbau und den mangelnden Tragekomfort der Brille hin (s. Tabelle 19).

Tabelle 18: Abschlussfragebogen (N = 29).

Item	Ja [%]	Nein [%]	Anmerkungen einzelner Probanden
Ließ sich die Kamera-Brille komfortabel am Kopf aufsetzen?	70.0	30.0	„drückt auf der Nase“; „drückt auf die Sehbrille“; „Helm kann nicht in gewohnter Position getragen werden“
Ließ sich die Kamera-Brille komfortabel am Kopf absetzen?	90.0	10.0	„drückt“; „Kabel hängt“
Empfanden Sie die Fassung der Kamera-Brille als stabil?	90.0	10.0	„Nasenbügel unstabil“; „rutscht von der Nase“
War die Kamera-Brille im Bereich der Nasenauflage komfortabel?	40.0	60.0	„Bügel störend“; „Nasenbügel zu klein und drückt“; „verengte Luftzufuhr“
War die Kamera-Brille im Bereich der Stirnauflage komfortabel?	96.7	3.3	--
War die Kamera-Brille im Bereich der Auflage hinter den Ohren komfortabel?	96.7	3.3	„Helm sitzt darüber“; „starkes Druckgefühl durch Ohrmuschel“;
Verrutschte die Kamera-Brille unter dem Helm, wenn Sie den Kopf bewegt haben?	6.7	93.3	„bei Kopfdrehungen“
Hatten Sie beim Tragen der Kamera-Brille Missempfindungen (z. B. Drücken) unter dem Helm?	23.3	76.7	„Nasentrücken“, „Ohren“, „Stirn“
Hatten Sie beim Tragen eine eingeschränkte Sicht nach außen durch die Kamera-Brille?	36.7	63.3	„ein Auge behindert“; „Ablenkung“; „Gestell im Blickfeld“; „ungewohnt“
Hatten Sie beim Tragen eine eingeschränkte Sicht auf die Instrumente durch die Kamera-Brille?	20.0	80.0	„beim Anflug“; „Crosscheck erschwert“; „Kopf musste öfter gedreht werden als sonst“
Hat Sie die Kamera-Brille beim Tragen an einer sicheren Flugführung gehindert?	16.7	83.3	„bestimmte Bereiche nicht einsehbar“; „fühlte mich nicht wohl – Prüfungssituation“; „Reizung am Nasentrücken“; „mehr Kopfbewegungen als sonst nötig“
Spüren Sie im Nachhinein Druckstellen durch die Kamera-Brille?	37.9	62.1	„Nasentrücken“; „30 Min. sind ok“
Spüren Sie im Nachhinein Juckreiz durch die Kamera-Brille?	6.9	93.1	„Nase“
Spüren Sie im Nachhinein Schmerzen durch die Kamera-Brille?	3.4	96.6	„Nasentrücken“
Hat Sie die Verkabelung mit dem Eye Tracking System bei der sicheren Flugführung gestört?	0.0	100.0	--
Hat Sie die Anwesenheit der Versuchsleiterin im Simulator bei einer sicheren Flugführung gestört?	0.0	100.0	--
Waren Symptome von Simulator Sickness durch das Eye Tracking System verstärkt ausgeprägt?	0.0	100.0	--

Knapp zwei Drittel der Stichprobe konnten sich vorstellen, dass Eye Tracking im Realflug anwendbar ist. Einzelne Fluglehrer gaben aber die Ablenkung des Flugschülers und die Einschränkung der Sicht zu bedenken. Schlussfolgernd lässt sich ableiten, dass unter Betrachtung des Stands der Technik des Dikablis Systems noch flugsicherheitsbezogener und ergonomischer Verbesserungsbedarf besteht, bevor es für den Realflug sicherheitsunkritisch einsatzfähig ist. Rein deskriptiv überwog bei der Bewertung von Eye Tracking für eine Eignung im Realflug der Anteil der Flugschüler, welcher die Methode als zweckmäßig bewertete (Abbildung 60).

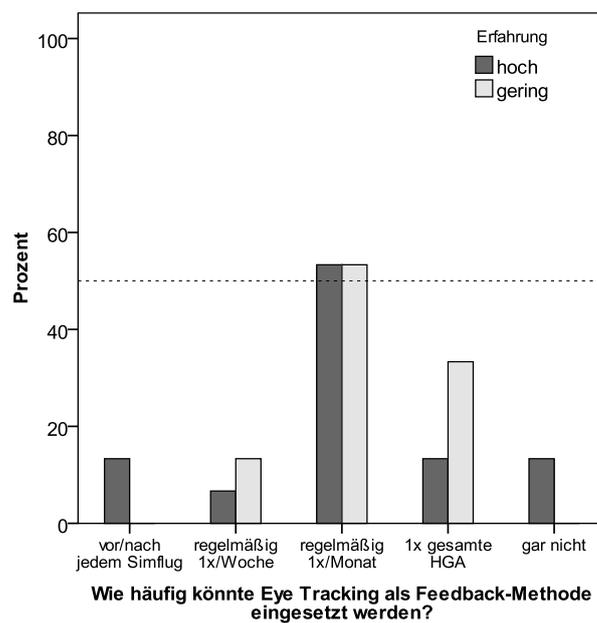


Abbildung 59: Eye Tracking als Feedback-Methode (N = 29).

Tabelle 19: Nützlichkeitsevaluation.

Item	Antwort	[%]	Anmerkungen einzelner Probanden
Wie häufig könnte Eye Tracking als Feedback-Methode eingesetzt werden?	Vor/nach jedem Simulatorflug	6.7	„Verbesserung des Gerätes (Befestigung am Helm!!!) notwendig“; „draußen gibt es auch keine Brille, train as you fly“; „Problem Zeitfaktor“; „die Methode sollte für jeden Ausbildungsabschnitt genutzt werden“; „derzeit zu unangenehmes und störendes Druckgefühl am Nasenbein“
	Regelmäßig einmal pro Woche	10.0	
	Regelmäßig einmal pro Monat	53.3	
	Einmal in der gesamten HGA	23.3	
	Gar nicht	6.7	
Abgesehen vom technischen Stand des Eye Tracking Systems (Verkabelung); eignet sich Eye Tracking als Feedback-Methode beim Realflug?	Ja	61.5	„erprobungswürdig“; „Flugsicherheit nicht sichergestellt“; „geringer Nutzen“; „im Simulator würde reichen“; „könnte Flugschüler ablenken“; „nur, wenn eingeschränkte Sicht auf die Instrumente behoben wird“
	Nein	38.5	

Anmerkung: HGA = Hubschrauberführergrundausbildung.

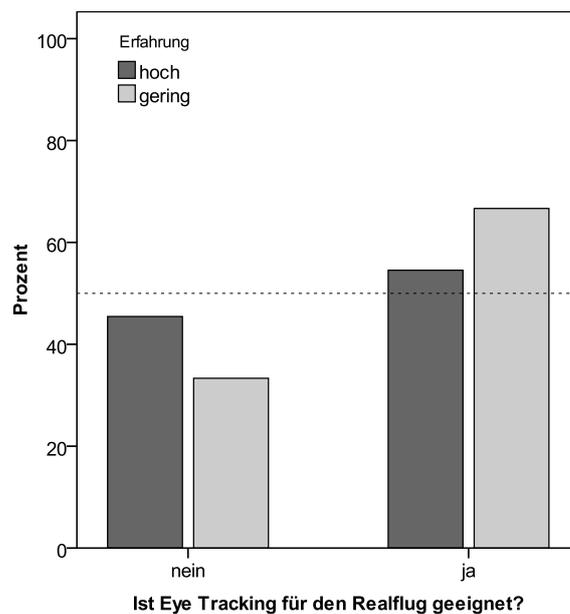


Abbildung 60: Eignung von Eye Tracking für den Realflug.

4.2.4 Zusammenfassung des Hauptversuchs

Nach Festlegung eines Cutoff-Wertes für Target Fixations von 3000 ms zeigte die inferenzstatistische Auswertung, dass sich in Konformität mit den Hypothesen ein signifikanter Interaktionseffekt für verschiedene Erfahrungs- und Anforderungsgruppen finden lässt: Während Flugschüler bei der Mission mit höheren visuellen Anforderungen (Gruppenversorger) mehr zu Target Fixations neigten, tätigte die Gruppe der Fluglehrer bei der Mission mit geringeren visuellen Anforderungen (Pinnacle) mehr Blickfixierungen. Die Nachfolgeanalyse anhand von Flugphasen deckte die unterschiedlichen Scanning Techniken der Untersuchungsgruppen auf und indizierte unter anderem, dass Fluglehrer über den Einsatz von Target Fixations von ihrem geschulteren peripheren Sichtfeld profitieren sowie vorhandene Informationskanäle effektiver nutzen als Flugschüler. Die Güte des Modells lag bei 54 % und war damit befriedigend. Bei Änderungen im Schwellenwert für Target Fixations wurde keine relevante statistische Verbesserung der Aufklärungsgüte erreicht.

Nach Analyse der Abweichung objektiver und subjektiver Daten zeigte sich, dass die Piloten sich hinsichtlich ihrer Scanning Techniken zum Teil verschätzen; beispielsweise überschätzen sie den Anteil von Blicken in die Außenwelt und unterbewerten den Anteil der Blicke auf die Instrumente. Dies gilt nicht nur für unerfahrenere Luftfahrzeugführer, sondern teilweise auch für erfahrene Piloten. Für letztere fällt die mittlere Selbsteinschätzung dennoch realistischer aus als für Flugschüler. In diesem Zusammenhang ergab sich ein signifikanter, linear-positiver Zusammenhang zwischen einer korrekten Selbsteinschätzung der Scanning Techniken und der Performance beim Landemanöver: Subjektiv besonders guten Performern sind die getätigten Blicke in Außenwelt oder Cockpit signifikant bewusster als subjektiv schlechteren.

Bei der ergonomischen Bewertung der Eye Tracking Methode stellten sich die Einschränkung der Sicht und Druckstellen auf dem Nasenrücken bei und nach Tragen der Kamerabrille als problematisch heraus. Der subjektive Komfort bei Anwendung von Eye Tracking wurde von den Probanden dennoch mit „akzeptabel“ eingestuft und die große Mehrheit fühlte sich an einer sicheren Flugführung nicht gehindert. Die Piloten konnten sich größtenteils vorstellen, Eye Tracking einmal pro Monat in der Heeresfliegerwaffenschule anzuwenden und gaben zu zwei Dritteln an, dass die Methode im Realflug einsatzfähig ist.

5 Diskussion

5.1 Scanning Techniken von Hubschrauberpiloten

Hauptanliegen dieser Arbeit war die Bereitstellung eines Algorithmus zur Analyse der Scanning Techniken von Hubschrauberpiloten. Aufzuklären waren hiermit Unterschiede in Scanning Techniken zwischen erfahrenen und unerfahrenen Piloten. Darüber hinaus sollte die Nützlichkeit der Eye Tracking Methode bewertet und anhand wissenschaftlicher Gütekriterien reflektiert werden.

Zur Analyse der Scanning Techniken wurden Target Fixations als Indikator der Aufmerksamkeit und visuellen Informationsakquise herangezogen und in Anlehnung an Empfehlungen der Luftfahrtbehörden sowie anhand eines statistischen Berechnungsalgorithmus operationalisiert. Die Hypothesen der Untersuchung bezogen sich auf ein multivariates Modell, innerhalb dessen der Einfluss von Flugerfahrung und visuellen Anforderungen der Mission auf Scanning Techniken, subjektive Beanspruchung und Leistung geprüft wurde.

Die Ergebnisse der Eye Tracking Studie legen nahe, dass – in Konformität mit der Interaktionshypothese – ein **wechselseitiger** Einfluss zwischen Flugerfahrung und visuellen Anforderungen der Mission auf Scanning Techniken von Hubschrauberpiloten existiert. Mithin unterscheidet sich das Auftreten von Target Fixations zwischen Hubschrauberpiloten mit unterschiedlicher Expertise in den hier untersuchten Landemanövern. Gezeigt werden konnte, dass insgesamt bis zu **10 %** der Blicke **Target Fixations** sind, deren Einsatz sich wie folgt gestaltet:

Landung im simulierten Gelände: Im Vergleich zu unerfahrenen überwiegen bei erfahrenen Piloten Target Fixations in die Außenwelt in Phasen höherer Beanspruchung (Start/Landung), wodurch sie sich vermutlich die Referenzinformation des Geländes effektiver zu Nutze machen. Unerfahrenere Piloten neigen insbesondere in der Überflugphase im Gegensatz zu Experten zu Target Fixations auf die Instrumente. Offenbar profitieren Experten im Geländeflug von ihrem größeren parafovealen und peripheren Sichtfeld (vgl. Williams, 1995), was durch den höheren Gesamtanteil an Target Fixations in dieser gegenüber der unerfahrenen Gruppe angedeutet wird. Unerfahrene Piloten könnten in diesem Fall lernen, mehr Target Fixations bei Start und Landung im richtigen Moment auf Referenzobjekte zu richten und den Instrumenten Check im Überflug kürzer durchzuführen. Bei der Interpretation von Target Fixations können aber andere Ursachen nicht eindeutig ausgeschlossen werden. Hierunter fallen beispielsweise die Tendenz zu Monotonieempfinden oder das Schonen von Aufmerksamkeitsressourcen bei erfahrenen Piloten.

Landung über simuliertem Gewässer: Im Vergleich zu unerfahrenen treten bei erfahrenen Piloten Target Fixations auf die Instrumente vermehrt in der Start- und Überflugphase auf. Aufgrund mangelnder Referenzinformation stabilisieren sie so scheinbar frühzeitig das Luftfahrzeug, während Flugschüler eher zu Target Fixations in die Außenwelt neigen. Fluglehrer orientieren sich im Überflug möglicherweise frühzeitiger am Gruppenversorger (Landepunkt) als Flugschüler. Fazit: Experten nutzen beim Flug über Wasser verhältnismäßig mehr Informationen aus den Instrumenten und führen insgesamt weniger Target Fixations durch als unerfahrenere Piloten. Flugschüler können hier lernen, Target Fixations sparsamer einzusetzen und bei Start und Landung mehr auf die Instrumente zurückzugreifen.

Der gefundene Interaktionseffekt zu Scanning Techniken von Hubschrauberführern ist deswegen bedeutsam, weil er die bisher eher inkonsistente Befundlage erklären könnte: Bei steigender Beanspruchung durch eine Aufgabe fanden Forscher bisher entweder Visual Tunneling oder Visual Scanning Randomness (z. B. Colvin et al., 2003; Velichkovsky et al., 2003; DiNocera et al., 2007; Tsai et al., 2007; Metz, 2009). Verknüpft man bisherige Befunde und die Ergebnisse dieser Untersuchung, lassen sich die empirischen Inkonsistenzen eventuell dadurch rationalisieren, dass Flugerfahrung und Anforderungen im multivariaten Ansatz nur als Faktorkombination zur Erklärung von variierenden Blicktechniken herangezogen werden können. Dies wird auch im Vergleich zu den Ergebnissen von Sullivan et al. (2011) deutlich: Die Autoren fanden für Flugerfahrung allein nur 42 % Varianzaufklärung der Blickdauer; die Kombination mit Aufgabenanforderungen erbrachte in der vorliegenden Studie immerhin 54 % (Target Fixations: $\eta^2 = .12$; mittelstarker Effekt).

In Konkordanz mit der Studie von Dixon et al. (1990), bei der es darum ging, wie schnell C 130 Piloten (Flächenflugzeug) hinsichtlich ihrer Scanning Techniken an ein kleineres Sichtfeld adaptieren, kann auch hier eine effektivere Scanning Technik von Experten bei einer visuell anspruchsvolleren Mission angenommen werden (zeitlich frühere Nutzung der Instrumente und Target Fixations ins Cockpit). Die Befunde sind außerdem mit der neueren Studie von Sullivan et al. (2011) konform, wo eine Navigationsaufgabe durch erfahrenere Hubschrauberpiloten (H-60 Navy Helicopter) besser bewältigt wurde, wenn diese eher ins Cockpit, anstatt in die Außenwelt blickten. Nichtsdestotrotz ist der direkte Vergleich der Studien untereinander mit Vorsicht vorzunehmen, da Scanning Techniken immer im spezifischen Kontext von Luftfahrzeugmuster und Aufgabe im Experiment zu interpretieren sind (s. ausführlich Abschnitt 5.4.1).

5.2 Nützlichkeit der Eye Tracking Methode

Die vorliegende Eye Tracking Studie hat gezeigt, dass sich beim Vergleich zwischen gemessenen Scanning Techniken und subjektiven Probandenaussagen (Interviews) die **Korrektheit der Selbsteinschätzung** abbilden lässt.

Die Teilnehmer der Studie überschätzten im Mittel den Anteil ihrer Blicke nach außen und unterschätzten im Mittel ihre Instrument Checks. Diese Fehleinschätzungen waren für Flugschüler erwartungsgemäß stärker ausgeprägt (z. B. absolute Verschätzung beim Instrumenten Check: Fluglehrer = 10 %; Flugschüler = 26 %).

Damit wird die Erkenntnis aus anderen Eye Tracking Studien unterstützt, wonach Versuchspersonen nicht immer in der Lage sind, ihre Scanning Techniken richtig einzuschätzen (z. B. Automobil: Sprenger, 2008; Medienforschung: Geise, 2012). Geise (ebd., S. 47) erklärt: *„Zu beachten ist [...], dass zwischen den Angaben im Rahmen der Befragung und der über Eyetracking gemessenen tatsächlichen visuellen Informationsaufnahme erhebliche Unterschiede bestehen können. So kommt es beispielsweise nicht selten vor, dass der Proband einzelne Bereiche des Stimulus intensiv foveal fixiert hat – sich aber im Fragebogen nicht daran erinnern kann“*. Neben der Unfähigkeit, sich präzise zu erinnern, kommt auch die Tatsache zum Tragen, dass die Scanning Techniken der Hubschrauberführer noch nie vergleichbar im Fokus standen bzw. gemessen wurden, entsprechend keine Referenzinformation für die Probanden existierte und somit die Selbsteinschätzung schwierig war. Eine andere Perspektive liefert die Betrachtung des Leistungsanspruchs, wonach ein Hubschrauber grundsätzlich anhand der Außenwelt zu führen ist, der damit verbundenen Erwartungen der Piloten hinsichtlich ihrer Scanning Techniken und somit sozial erwünschter Antworttendenzen. *„Problematisch ist [...], dass einige Probanden dazu tendieren, ihr Handeln nachträglich zu plausibilisieren und zu rationalisieren; das gilt insbesondere – vermutlich aus Gründen sozialer Erwünschtheit – wenn der Blickverlauf bestimmte Verzögerungen oder Schwierigkeiten beim Lösen der Aufgaben [offenbaren könnte]“* (Geise, 2012, S. 56; vgl. auch Bierhoff & Frey, 2011).

Eine korrekte Selbsteinschätzung der eigenen Scanning Techniken ist deswegen in der Luftfahrt nicht trivial, weil diese entscheidend für die sichere Flugführung sind und zur Vermeidung von Aufmerksamkeitsfehlern eingesetzt werden sollten. Möglicherweise kommt es dabei nicht auf die Einhaltung einer standardisierten Empfehlung an (vgl. EASA, 2010), sondern auf eine adäquate Reflektion der individuellen Scanning Technik:

Piloten, die ihre Scanning Techniken realistischer einschätzen, zeigen in dieser Untersuchung subjektiv die präzisere Landung ($r = .31$, $p = .035$). Besonders guten Performern sind ihre Scanning Techniken signifikant bewusster als schlechteren (gemessen an der Abweichung objektiver und subjektiver Eye Tracking Daten sowie subjektiver Leistungsbewertungen). Die subjektive Performance steht jedoch nicht mit dem durch Eye Tracking gemessenen Anteil an Target Fixations in Verbindung. Da die Berechnung von Target Fixations auf den Hinweisen der EASA (2010) basierte, liegt die Vermutung nahe, dass die Empfehlungen nicht unbedingt eine Voraussetzung dafür darstellen, einen Hubschrauber sicher zu landen.

Sicher ist die Selbsteinschätzung der Scanning Techniken dabei nicht die einzige Determinante der Leistung, aber die präzise visuelle Informationsakquise bzw. die Metakognition hierüber kann zentrale Vorstufe für Entscheidungsfindung und Handeln in der Luftfahrt sein (EASA, 2010). Die Ergebnisse gehen über die von Sullivan et al. (2011) hinaus; hier wurde zunächst nur festgestellt, dass objektive Blickparameter allein nicht direkt mit der Leistung bei einer Navigationsaufgabe korrelieren. Die Selbsteinschätzung als Determinante wurde bisher nicht geprüft. Zu bedenken ist aber, dass bei Sullivan et al. (ebd.) objektive Leistungsparameter (Abweichung von einer Flugroute) verwendet wurden, die Ergebnisse dahingehend also nur **mit Vorsicht** verglichen werden können.

Die Reflektion über eigene Scanning Techniken kann mithilfe von Eye Tracking erleichtert werden. Damit können nicht nur unerfahrene Piloten schneller lernen, ihre Scanning Techniken gezielt einzusetzen, sondern es können auch im weiterführenden Training von erfahrenen Piloten Debriefings nach Flügen hiermit angereichert werden. Salas, Wilson, Burke, Wightman und Howse (2006, S. 52) merken hierzu an: *“Providing trainees with the opportunity to demonstrate mastery of the learned knowledge, skills, and attitudes enables them to apply what they have learned, receive feedback, and correct their performance before reaching the operational environment”*. Den Gedanken in Bezug auf Eye Tracking weiterführend, ist das Verfahren als Rückmeldewerkzeug aber vorerst nur für das **Simulatortraining** denkbar, denn die Tatsachen, dass die Kamerabrille für einen nicht vernachlässigbaren Teil der Piloten Einschränkungen der Sicht nach außen und auf die Instrumente sowie Druckstellen auf dem Nasenrücken bewirkt, bilden zentrale **Intrusivitätsfaktoren**, die die Unbedenklichkeit der Anwendung von Eye Tracking im Cockpit (insbesondere im Realflug) einschränken.

Die Ergebnisse aus der ergonomischen Bewertung des Eye Tracking Systems zeigen auch, dass sich über 80 % der Stichprobe durch Eye Tracking nicht an einer sicheren Flugführung gehindert fühlen und der Komfort der Eye Tracking Brille im Durchschnitt mit akzeptabel bewertet wird. Die Piloten befinden mehrheitlich,

dass Eye Tracking einmal im Monat während der Ausbildung nützlich sein kann und können sich zu zwei Dritteln einen potenziellen Einsatz im Echtflug vorstellen.

5.3 Wissenschaftlichkeit der Eye Tracking Methode

Eye Tracking als wissenschaftliche Methode sollte den Gütekriterien der Sensitivität, Validität, Diagnostizität, Reliabilität, Nicht-Intrusivität, Ökonomie und Augenscheinvalidität standhalten. Wissenschaftliche Methoden sind **sensitiv**, wenn sie feine Schwankungen in gemessenen Größen abbilden können (Rubio et al., 2004). Spezifiziert auf den vorliegenden Anwendungsfall bedeutet dies: Ein Eye Tracking System sollte interindividuelle sowie aufgabenspezifische Auswirkungen auf das Auftreten von Target Fixations reflektieren können. Singuläre Haupteffekte auf Flugerfahrung oder den Anforderungen durch die Mission allein lassen sich hier zwar nicht aufdecken, aber gemäß der multivariaten varianzanalytischen Auswertung tritt ein signifikanter, mittelstarker Interaktionseffekt der Faktoren auf. Entsprechend wurde mithilfe von Eye Tracking Daten ein Unterschied für Target Fixations identifizierbar gemacht, der wahrscheinlich ($p = .033$) auf der Kombination von Erfahrung und Aufgabenanforderungen basiert. Somit erfüllt die Eye Tracking Methode den Sensitivitätsanspruch dieser Untersuchung, wonach interindividuelle und aufgabenbezogene Differenzen messbar sein sollen.

Das Eye Tracking Verfahren wäre dann **diagnostisch**, wenn es erlaubte die Faktoren zu identifizieren, auf die sich die gemessenen Unterschiede für Target Fixations zurückführen lassen. Schultheis (2004, S. 19) merkt hierzu an: „[ein diagnostisches Verfahren hat] aber den Nachteil, dass es nicht sehr sensitiv ist. [...] eine hoch diagnostische Methode [liefert] also keine Informationen über Belastungszustände, die durch andere Aspekte [als die diagnostizierten] verursacht sind. Insofern kann ein Maß nicht gleichzeitig hoch sensitiv und hoch diagnostisch sein“ (vgl. hierzu auch Amelang & Schmidt-Atzert, 2006). Gemäß dieser Erklärung und in Anlehnung an Geise (2012) kann über die Ursachen der aufgedeckten Unterschiede für Target Fixations (z. B. Einfluss des peripheren Sichtfelds) nur diskutiert werden, sie sind aber nicht direkt messbar.

Die **Validität** einer wissenschaftlichen Methode gibt an, inwiefern das Verfahren tatsächlich „nur“ auf die Veränderungen der gemessenen Größe reagiert bzw. inwiefern auch konfundierte Variablen oder Störgrößen in die Messung einfließen. Das für diese Studie herangezogene Modell weist eine befriedigende Varianzaufklärung von 54 % auf. Damit bleibt ein Teil der Varianz unerklärt, was aber insofern nicht überraschend ist, als dass die hier gemessene Größe auf eine bisher nicht vergleichbare Weise aus Blickdaten abgeleitet wurde und weitere Determinanten nicht ausgeschlossen werden können. Da es bislang keine, mit Eye Tra-

cking direkt vergleichbare bzw. diametrale Methode gibt, kann hier keine Aussage über die konvergente oder diskriminante Validität der Methode gemacht werden.

Die **Reliabilität** einer Methode meint das Ausmaß, inwiefern sich unter gleichen Bedingungen gleiche Messwerte erzielen lassen. Schultheis (2004, S. 20) gibt hierzu an: „Häufig wird die Reliabilität einer Methode informal daraus abgeleitet, dass sich in verschiedenen Studien, die dasselbe Maß verwendet haben, konsistent ein Zusammenhang zwischen dem Messwert und der (vermuteten) kognitiven Belastung zeigt“. Bisher wurden Target Fixations mit Eye Tracking nicht vergleichbar untersucht. Obwohl Studien zum Visual Tunneling Effekt existieren (Colvin et al., 2003; Velichkovsky et al., 2003; Metz, 2009), wurde hier häufig nur aus der prozentualen Nutzungshäufigkeit auf eine Blickfixierung geschlossen. Zur Reliabilitätsbewertung von Eye Tracking anhand der Messung von Target Fixations bedarf es weiterer Studien, die Effekte für dieses Maß in variierenden Untersuchungskontexten analysieren (z. B. im Realflug, im Vergleich von Nicht-Piloten und Piloten oder für andere Anwendungsfälle wie beispielsweise Operateure in Kernkraftwerken beim Abarbeiten von Checklisten).

Wissenschaftliche Methoden sind als **intrusiv** zu bewerten, wenn sie die eigentliche Aufgabe des Operateurs stören. Die Intrusivität der Eye Tracking Methode mit dem Head Mounted Eye Tracking System Dikablis Cable ist, insbesondere für die Anwendung des Verfahrens im Realflug kritisch zu bewerten. Für die Anwendung in einer simulierten Arbeitsumgebung ist das Ausmaß der Intrusivität akzeptabel, erhöht sich aber auf ein inakzeptables Niveau, wenn Versuche länger als etwa eine Stunde pro Proband dauern und das Tragen eines Helmes erfordern. Hier wäre Hardware wünschenswert, die sich außen am Pilotenhelm befestigen ließe.

Die **Ökonomie** einer Methode erschließt sich aus dem Verhältnis von Kosten und Nutzen des Verfahrens. Zu den Kosten zählen im vorliegenden Fall nicht nur der finanzielle Anschaffungsaufwand eines Eye Tracking Systems, sondern auch die Einarbeitungszeit in die Hardware-Komponenten, die notwendigen Auswertungs-Softwares und die Erfordernisse, die durch den Versuchsaufbau entstehen. Das Eye Tracking System Dikablis Cable der Firma Ergoneers GmbH kostet in der hier eingesetzten Version inklusive Aufnahmegerät ca. 25.000,- Euro⁴¹ und liegt damit in einem für physiologische Messgeräte bzw. andere Eye Tracking Systeme (z. B. Tobii Glasses System ca. 24.000,- Euro⁴²) durchschnittlichen Bereich. Die Einarbeitung in das System ist kaum zu bemessen, sie beträgt bei täglicher (!) Auseinandersetzung mit dem Verfahren und explorativen Vorversuchen mit dem

⁴¹ Angebot Ergoneers GmbH, 21.04.2009.

⁴² Tobii Technology, Online-Kommunikation vom 12.06.2012.

Ziel der Methodenkompetenz insgesamt etwa 200 Stunden. Um die Eye Tracking Daten in der hier durchgeführten Weise auszuwerten, bedarf es einer Anzahl von sieben verschiedenen Softwares. Auch dies ist unter dem Aspekt der zeitnahen und möglichst effizienten Ergebnismeldung an den Operateur relativ aufwändig. Der Aufwand, der durch den Versuchsaufbau entsteht, ist als relativ gering einzustufen bzw. vergleichbar mit anderen Versuchsdesigns mit physiologischem Hintergrund. Als größter Zeitfaktor fließt neben der eigentlichen Messung die Kalibrierung des Systems ein, welche bei gründlicher Durchführung mit etwa 20 bis 30 Minuten pro Proband zu kalkulieren ist.

Die genannten Kosten der Methode relativieren sich insbesondere in Kombination mit ihrer Intrusivität bei längeren Messstrecken nur bedingt am Nutzen des Eye Tracking Verfahrens zur Analyse von Scanning Techniken im Cockpit, aber die Methode ist die derzeit einzige, die die gezielte Analyse und entsprechende Rückmeldung für Hubschrauberpiloten ermöglicht.

Augenscheinvalidität liegt vor, wenn ein wissenschaftliches Verfahren für die Probanden einer Studie, die meistens auch die fokussierten Operateure sind, als sinnvoll erachtet wird. Wie Sullivan et al. (2011, S. 871) anmerken, stehen Fluglehrer häufig vor dem Dilemma, wann und wie Flugschülern die beste Rückmeldung vermittelt werden kann. Die Autoren erklären: *„Busy flying in a challenging environment, avoiding terrain and obstacles, the instructor has few opportunities to understand what mistakes were made by the student, when the mistakes were made, and how to explain those mistakes to the student to provide an opportunity to learn from them, all within a few minutes of the mistake being made“*. Dass Eye Tracking Daten genau diese Rückmeldemöglichkeit liefern können, wurde in dieser Studie angezeigt (vgl. Sullivan et al., 2011) und wird außerdem durch das hier dargestellte Ergebnis untermauert, wonach sich knapp zwei Drittel der Piloten vorstellen können, dass Eye Tracking einmal im Monat im Simulatortraining zum Einsatz kommt. Eine zusammenfassende Bewertung der wissenschaftlichen Güte von Eye Tracking in Bezug auf die Messung von Target Fixations liefert Tabelle 20.

Tabelle 20: Zusammenfassende Bewertung.

Sensitivität	Diagnostizität	Validität	Reliabilität	Intrusivität	Ökonomie	Augenscheinvalidität
++	--	+	?	-	-	++

Anmerkung: ++ = Kriterium erfüllt, + = Kriterium mit Einschränkungen erfüllt, - = Kriterium kritisch, -- = Kriterium nicht erfüllt, ? = ungeklärt.

5.4 Generalisierbarkeit und Einschränkungen der Studie

5.4.1 Zur Rolle von Target Fixations nach Rücksprache mit Experten

Die Befunde der Untersuchung fußen auf der Annahme, dass Target Fixations von Experten intendiert und gezielt sind. Target Fixations, die bei unerfahrenen Piloten auftreten, werden hingegen teilweise eher dysfunktional und ineffektiv bewertet (Pre-Test-Experteninterviews, N = 6 am 17.10.2011). Dieser Unterschied stellt aber nur eine mögliche Interpretation des hier gefundenen Interaktionseffekts dar. Daher wurden **alternative Interpretationen** des Effekts in Erwägung gezogen. Während der Diskussion der Ergebnisse in einer erneuten Expertenrunde (Post-Test, N = 3 am 12.09.2012) wurde beispielsweise debattiert, ob Target Fixations aufgrund von **Monotonie oder Unterforderung** während der Untersuchung aufgetreten sein können (vgl. Joiko, Schmauder, & Wolff, 2012; Buld, 2000; DeWaard, 1996; Johannsen, 1979). Die erfahrenen Piloten schlossen Ermüdung zwar aus (s. auch Befunde von Previc et al., 2009, wonach Scanning Techniken nicht durch Ermüdung beeinflusst werden sowie, dass M = 0 beim Item „Ermüdung“ aus dem Simulator Sickness Questionnaire) und betonten die Rolle von Target Fixations beim Auffinden der optimalen Fluglinie und der Bestimmung sowie Einhaltung des Landewinkels („*knowing what is needed*“, Wickens, 1992, zitiert nach Uhlarik & Comerford, 2002, S. 2). Dies rechtfertigt laut Expertenaussagen auch das Target Fixation Maximum bei Fluglehrern im Geländeflug von 25 Sekunden. Berücksichtigt man allerdings den **Einfluss von Vorwissen**, das heißt der schematischen Wahrnehmung von **missionstypischen Reizen** durch Experten (vgl. Henderson, 2003: schematisches Szenenwissen), kann nicht ausgeschlossen werden, dass die Ergebnisse nur auf diesen Geländeflug im Simulator anwendbar sind. Denkbar ist hier, dass Experten Target Fixations deswegen einsetzen, weil sie eine **genaue Erwartung** bezogen auf die Szene haben (z. B. „*Jetzt kommt der Baum an der Ecke..., jetzt erscheint das Haus mit dem roten Dach...*“). Damit muss die mögliche Messung von **Artefakten** im Hinblick auf Target Fixations bedacht werden. Die Berechnung basierte auf Empfehlungen der Luftfahrtbehörden (EASA, 2010) zum optimalen Scanning, welche zunächst für alle Luftfahrzeugführer gelten: „*It should be of interest to all pilots, regardless of the type of aircraft they fly*“ (S. 3). Es bleibt aber **fraglich**, ob die Blicktechniken von zivilen Airline-Piloten bei Reiseflughöhe (ca. 35000 ft) mit denen von Hubschrauberpiloten bei ca. 300 ft **vergleichbar** sind. Darüber hinaus kann die Geländenavigation bzw. die Orientierung über größeren Gewässern per se eine längere Blickdauer erfordern. Diesbezüglich war in den vorliegenden Ergebnissen ein gravierender Unterschied der Blickdauer der Piloten im Vergleich zu den Hubschrauberpiloten bei Sullivan et al. (2011) evident: In der aktuellen Studie mit Navigationsaufgabe von der Gruppe

um Sullivan betrug die mittlere Blickdauer OTW = 231 ms. Die Piloten der vorliegenden Untersuchung sahen im Mittel sechs Mal so lang (M = 1360 ms) in den Luftraum und immerhin noch drei Mal so lang auf die Instrumente als die Hubschrauberführer bei Sullivan et al. (ebd.), die eine Karte im Simulatorcockpit mit M = 271 ms fixierten. Daneben waren die Blickverteilungen in beiden Studien aber gleich: Bei Sullivan et al. (ebd.) wurde zu 57.7 % OTW gescannt, hier wurde ein Anteil von 58.2 % OTW gefunden. Die Unterschiede in der Blickdauer könnten auf **technische Ungleichheiten** im Eye Tracking System (Fixation vs. Blick, s. Abschnitt 2.1.6) oder aber die differierende Instruktion der Piloten von Sullivan et al. zurückzuführen sein, welche eine Navigationsaufgabe neben der Flugführung bearbeiteten und dabei kein explizites Landemanöver durchführen mussten (Geländeflug ohne Simulatorbewegung bei 150 ft). Zu bedenken ist außerdem, dass Eye Tracking, wie bereits dargestellt, **nicht** in der Lage ist, **parafoveale Wahrnehmungsmuster** zu messen, sondern lediglich direkte Blicke codiert werden. Das Fehlen einer Zweitaufgabe (z. B. Navigation) und die Nichtberücksichtigung parafovealer Informationsverarbeitung, die beim Paradigma mit nur einer Primäraufgabe (Luftfahrzeug sicher landen) vielleicht zu einem nicht zu vernachlässigenden Anteil involviert ist, könnte die längere Blickdauer der hier inspizierten Piloten erklären, da diese vielleicht weniger Sakkaden bzw. Blickwechsel durchführen mussten als die Luftfahrzeugführer in der Studie von Sullivan et al. (2011).

Die Expertengruppe, welche sich mit den Ergebnissen auseinandersetzte, wies auf einen weiteren Faktor hin, der die stark abweichende Blickdauer in den verschiedenen Studien erklären kann: die **ergonomische Gestaltung** des Cockpits. Der Helicopter (Navy H-60) bei Sullivan et al. (2011) verfügt über das typische Glass Cockpit mit Multifunktionsdisplays, hat aber im Gegensatz zum EC 135 den ergonomischen Vorteil, dass die Anzeigen bei unauffälligem Systemstatus alle mit einem hellgrünen LED-Rahmen leuchten (s. Abbildung 61). Bei Auffälligkeiten bzw. Störungen im System wechseln die Farben auf amber bzw. rot. Daher kann der Pilot mit relativ kurzen Blicken ins Cockpit eine erschöpfende Instrumentenkontrolle durchführen. Im EC 135 liegt keine vergleichbare visuelle Unterstützung durch grüne LEDs vor, sodass eine längere Blickdauer ins Cockpit erforderlich sein kann. Der genannte Unterschied im Einfluss des ergonomischen Designs untermauert einmal mehr, dass die Ergebnisse von (Eye Tracking) Studien im Cockpit immer im Kontext der spezifischen, für die Untersuchung geschaffenen Bedingungen interpretiert werden müssen. Die **Generalisierbarkeit** von Eye Tracking Befunden ist daher **eingeschränkt** (vgl. Geise, 2012) und beläuft sich hier auf: Landemanöver unter VFR im EC 135 Hubschraubersimulator bei simuliertem Gelände (Pinnacle) und über simulierten Gewässern (Einsatzgruppenversorger). Folgestudien zu Target Fixations unter anderen Bedingungen sind entsprechend indiziert.



Abbildung 61: Cockpit Navy H-60⁴³ mit LED-Unterstützung.

5.4.2 Zur Rolle von Performance und Workload

Erwartungsgemäß haben Fluglehrer eine bessere **Leistung** in beiden Missionen als Flugschüler. Berücksichtigt werden sollte bei der Interpretation dieses Ergebnisses aber nicht nur der Einfluss der Flugerfahrung, sondern auch die Tatsache, dass es sich bei Fluglehrern um eine **Selbsteinschätzung** der Leistung handelte, während die Einschätzung der Schüler durch die Fluglehrer vorgenommen wurde. Damit lagen Unterschiede in der Lokation der Bewertung vor (Selbst- vs. Fremdeinschätzung), auf die ebenfalls Varianzanteile zurückzuführen sein können. Für zukünftige Untersuchungen wäre ein objektives Leistungskriterium (z. B. durch PC errechnete Abweichung vom Landepunkt) wünschenswert, um mögliche Verzerrungen durch subjektive Einflüsse zu vermeiden. Innerhalb der Hubschrauberführergrundausbildung wird die Leistung von Piloten aber bisher auf den hier angeführten Dimensionen (mentale Restkapazität, Abweichung vom Landepunkt, Airmanship) durch das Urteil von Fluglehrern vorgenommen; objektivere Kriterien werden nicht herangezogen. Entsprechend war es gerechtfertigt, die übliche Leistungsbewertung für die Eye Tracking Studie zu nutzen. Wie bereits Sullivan et al. (2011) resümieren, korrelieren auch in dieser Untersuchung Eye Tracking Daten nicht direkt mit Performance-Angaben der Piloten. Dies ist sicherlich auch der Tatsache geschuldet, dass übliche Leistungsmaße keine auf Scanning Techniken bezogenen Krite-

⁴³ http://media.defenseindustrydaily.com/images/ELEC_MH-60R_Trainer_Cockpit_lg.jpg [Zugriff: 28.04.2013].

rien enthalten. Denkbar ist deswegen in Zukunft, bei Leistungseinschätzungen Scanning Techniken zu berücksichtigen.

Die Subskala mentale Anforderungen des NASA-TLX variierte zwar signifikant zwischen den Erfahrungsgruppen, aber für die subjektive **Gesamt-Workload** der Piloten wurden trotz Erfahrungs- und Anforderungsvariation **keine relevanten Unterschiede** aufgedeckt. Dieses Ergebnis verwundert vor dem Hintergrund, dass die Missionen verschiedene visuelle Anforderungen hatten, deckt sich allerdings mit bisherigen Befunden zur Workload-Messung beim Simulatortraining der Heeresfliegerwaffenschule (Stein & Robinski, 2011a). Auf der einen Seite ist denkbar, dass die **mündliche Instruktion** des NASA-TLX für die Probanden **zu undeutlich** war (auch, wenn jeder Proband einzeln instruiert wurde). Auf der anderen Seite könnte, wie auch beim Performance-Maß, zugrunde liegen, dass der NASA-TLX für den Faktor, auf dem sich die Missionen unterschieden, **nicht sensitiv** genug ist. Für die Detektion von Unterschieden in der subjektiv empfundenen Beanspruchung hinsichtlich des Mittelwertes aus mentalen, physischen, zeitlichen, anstrengungs-, leistungs- und frustrationsbezogenen Faktoren bedurfte es eventuell einer anderen Konstellation von Versuchsbedingungen. Die Variation der Anforderungen auf der **visuellen Ebene** (schnelle Informationsaufnahme und -auswertung aus dem Cockpit erforderlich, große Datenmenge aus den Instrumenten bei gleichzeitig schwer einschätzbarer Umgebungsinformation zu berücksichtigen, kaum fixe Referenzpunkte im peripheren Sichtfeld usw.) ist möglicherweise zu spezifisch für ein allgemeines Beanspruchungsmaß: „[Es] kann [mit dem NASA-TLX] nicht exakt bestimmt werden, ob es sich um die Auslastung zentraler Informationsverarbeitung oder der visuellen Verarbeitung handelt“ (Buld, 2000, S. 36; vgl. auch Rubio et al., 2004). Diese Vermutung deckt sich mit Befunden aus der Literatur, wonach mit dem NASA-TLX **nicht immer Unterschiede** für die Beanspruchung durch verschiedene Aufgabenausprägungen **innerhalb einer Sinnesmodalität** (hier beispielsweise verschiedene visuelle Bedingungen) gefunden werden.

Die Sensitivität des NASA-TLX bei Anforderungsvariation innerhalb ein- und derselben Sinnesmodalität

→ Horberry et al. (2006) untersuchten die Fahrleistung und Beanspruchung bei Variation der Fahraufgabe (FAS vs. Handy-Telefonat) sowie der visuellen Umgebungsbedingungen (einfache vs. komplexe Umgebung). Während der NASA-TLX klar zwischen den Aufgaben trennte, wurde für Unterschiede in den visuellen Umgebungsbedingungen keine Signifikanz erzielt: „*Workload did not significantly differ between the simple and complex environment conditions, indicating that distraction coming from outside the vehicle did not, as perceived by participants, increase workload*“ (S. 189).

- Nelson et al. (1998) verglichen ein HMD und ein Dom-Display hinsichtlich der visuellen Entdeckungsleistung und Workload. Zusätzlich variierten sie das Ausmaß auditorischer (z. B. "NL = nonlocalized, 2D = two-dimensional, 3D = three-dimensional auditory information") und visueller ("ML = mechanically limited, SL = software-limited") Information. Die Autoren fanden zwar Haupteffekte für Workload zwischen den Modalitäten der Stimuluspräsentation (visuell vs. auditorisch) aber nicht innerhalb der Modalitäten: "[...]. In addition no significant differences [...] in workload ratings were revealed between the 2D and 3D conditions or the none and NL conditions. [...] the full, ML, and SL conditions were not significantly different from one another" (S. 5).
- Metzger und Parasuraman (2005) verglichen verschiedene automatisierte Assistenzsysteme (Variation der visuellen Informationsdarstellung) für Fluglotsen hinsichtlich Leistung und subjektiver Beanspruchung. Im Ergebnis zeigte sich, dass „The effect of automation on ratings of mental workload failed to reach significance [...]. The contrast between automated and manual conditions [...] was also nonsignificant [...]“ (S. 11).
- Brewster und Cryer (1999) untersuchten, ob auditorische Informationen (kein Geräusch vs. Basisgeräusch vs. erweitertes Geräusch) die Gebrauchstauglichkeit eines Mobiltelefons verbessern. Sie verglichen die Beanspruchung über alle drei Bedingungen und fanden: "[...] tests showed that both audio treatments significantly lowered workload over the silent buttons (but no difference was shown between the two audio treatments)" (S. 3).
- Billings (2008) erhob die subjektive Workload in vier Experimentalbedingungen: die Versuchsteilnehmer sollten ein simuliertes unbemanntes Luftfahrzeug mit einer Maus oder einem Joystick (haptische Variation) entweder ohne oder mit 500 ms zeitlicher Latenz (zeitliche Variation) steuern. Bezogen auf den NASA-TLX ergab sich: "[...] however, workload failed to be affected by input device or latency" (S. 2095).

In Anlehnung an die dargestellten Befunde ist auch für das vorliegende Untersuchungsergebnis bezüglich des NASA-TLX denkbar, dass gerade Unterschiede in visuellen Anforderungen einer Mission nicht hinreichend mit dem NASA-TLX abgebildet werden (vgl. Denning et al., 2003). Dies gilt offenbar sogar dann, wenn die Skalenbeschreibungen des NASA-TLX, wie im vorliegenden Fall, exemplarisch auf den Arbeitsplatz spezifiziert werden (hier: Simulator).

Eine andere Möglichkeit, weshalb Beanspruchungsunterschiede hier nicht durch den NASA-TLX angezeigt werden, besteht darin, dass die Varianz der Flugerfahrung zu gering war. Die Flugschüler dieser Studie hatten **bereits einen gewissen**

Anteil an Flug- und Simulatorerfahrung und waren entsprechend nicht als „Tabula Rasa“ einzustufen. Somit könnten beide Gruppen – also Fluglehrer und Flugschüler – das Verhältnis von geforderten und aufgewendeten Ressourcen als im Mittel gleich empfunden haben. Letztgenannter Gedanke gilt vielleicht besonders deswegen, weil Flugschüler (wie oben erwähnt) eben aufgrund von Generationseffekten keine „Simulatornovizen“ (= geforderte Ressourcen geringer) aber Fluglehrer den größeren Erfahrungsschatz mit den Anforderungen der Flugführung haben (= aufgewendete Ressourcen geringer).

5.4.3 Zum Versuchsaufbau

Eine zentrale Frage zur Generalisierbarkeit der Ergebnisse auf die Realsituation des Hubschrauberflugs ist die nach der Validität der Datenerhebung in einer simulierten Umgebung. Zunächst muss festgehalten werden, dass bei einem Feldversuch im Realluftfahrzeug viele Störfaktoren auftreten können, die im Gegensatz zur Situation im Simulator nicht kontrollierbar sind (z. B. Wetterverhältnisse, Missionsinhalte, Auftreten von sicherheitskritischen Zwischenfällen, technische Ausfälle usw.). Im Simulator hingegen können klare Versuchsbedingungen hergestellt und kontrolliert werden, die Versuche sind kostengünstiger und darüber hinaus weniger sicherheitsgefährdend. Im vorliegenden Fall war die Versuchsdurchführung im Simulator schon deswegen indiziert, weil das Eye Tracking System bisher **nicht für Realflüge erprobt oder zugelassen ist** und innerhalb der Bundeswehr noch in keinem Flug- oder Simulatorkontext eingesetzt wurde. Dass Parameter des Ressourcenmanagements, wie mentale Beanspruchung und Leistung im Flugsimulator mit denen im Realflug vergleichbar sind, zeigt eine Studie von Svensson et al. (1999). Dennoch bleibt es ein Fakt, dass valide Aussagen über Target Fixations im Realflug nur nach Einsatz von Eye Tracking im echten EC 135 Cockpit getätigt werden können. Inwiefern Scanning Techniken im Simulator von denen im Realflug abweichen, wurde beispielsweise durch Studien im Automobilkontext untersucht (z. B. Stahl & Kleinkes, 2005), wobei sich hier die Scanning Techniken der Fahrer als vergleichbar herausstellten.

Neben dem Generalisierbarkeitsproblem von Simulatorstudien existiert, wie bereits erwähnt, die Möglichkeit der Messung von Artefakten. Schandry (1989, S. 111) definiert ein Artefakt als ein „*vom Messfühler aufgefangenes Signal, das anderen Ursprungs ist als das zu messende Biosignal*“. Übertragen auf den vorliegenden Fall bestünden Simulatorartefakte aus Blicken, die über 3000 ms dauern, aber nicht auf visuelle Anforderungen der Mission oder Flugerfahrung zurückzuführen wären, sondern z. B. auf die visuelle Verarbeitung der **Diskrepanz zwischen Simulations- und Realumgebung**. Ausgehend davon, dass in einer simulierten

Umgebung bei Vorhandensein vieler Objekte die Abweichung von der Realität mehr zum Tragen kommt als bei Nichtvorhandensein von Referenzobjekten und eventuell eine **andere Art der Linsenakkommodation aufgrund der virtuellen Umgebung** nötig ist, könnten gerade Fluglehrer in einer solchen Simulator zu Abweichungen ihrer reellen Scanning Techniken neigen. Kombiniert man die Unterschiede der Umgebungen in den Missionen mit dem Gedanken, dass Flugschüler eher der „Generation Simulator“ angehören, ist es denkbar, dass letztere in der simulierten Umgebung einen realistischeren Eindruck haben und natürlichere Scanning Techniken zeigen. Diese Interpretation wird von Ergebnissen untermauert, wonach Fluglehrer der Heeresfliegerwaffenschule die visuellen Eigenschaften der Simulatoren eher kritisieren als Flugschüler und darüber hinaus häufiger unter der **Simulatorkrankheit** leiden (Stein & Robinski, 2012), was ein Indiz für deren geringere Adaption an die Simulationsumgebung sein kann. Auch aus anderen Forschungsbereichen (Fahrsimulatoren) sind diese Effekte bekannt: Ältere Probanden (56 Jahre +) klagten im Vergleich zu Jüngeren (bis 35 Jahre) eher über mangelnden Komfort im Simulator sowie über die grafische Qualität, die Verzögerung von Steuereingaben und leiden eher unter Symptomen der Simulatorkrankheit. Nachweislich spielen ältere Personen weniger häufig Video- oder Computerspiele (Liu, Watson, & Miyazaki 1999), was in Kombination mit einem hohen Erfahrungsanteil mit der Realaufgabe (z. B. einen Pkw oder ein Luftfahrzeug steuern) ebenfalls für eine geringere Familiarität dieser Generation mit Simulationsumgebungen spricht. Hinzu kommt, dass die Flugschüler dieser Studie wie erwähnt in gewissem Sinne an Flüge im EC 135 Simulator adaptiert waren.

Nicht vollständig ausgeschlossen werden können außerdem **Reaktivitäts- und Transfereffekte** innerhalb der Stichprobe. Da sich die Datenerhebung über mehrere Tage erstreckte, ist es möglich, dass sich sowohl Missionsinhalte als auch Intention der Eye Tracking Studie („Analyse von Target Fixations“) unter den Piloten herum gesprochen haben und diese ihre Scanning Techniken im Rahmen der Messungen modifizierten. Auffällig waren die Aussagen einiger Piloten, wonach sie sich „wie in einer Prüfungssituation“ vorkamen (Flugschüler) oder „lange nicht bewertet wurden“ (Fluglehrer). Ausgehend davon, dass sich auch Merkmale wie soziale Erwünschtheit, Reaktivität und Bewertungsangst in einer Stichprobe normalverteilen, die Probanden randomisiert auf die Versuchsbedingungen verteilt und ausführlich über Freiwilligkeit und Anonymität aufgeklärt wurden, schränken diese Effekte die Validität der Untersuchung in eher geringem Maße ein.

5.4.4 Zum Eye Tracking System Dikablis Cable

Neben den oben diskutierten Komforteinschränkungen der Hardware des Eye Tracking Systems sollen an dieser Stelle software- und auswertungsbezogene Aspekte beleuchtet werden. Leider sind mit der hier verwendeten Version des Eye Tracking Systems **keine Sakkaden oder Blickwechsel** sowie **keine Pupillendilatationen oder Lidschläge** auswertbar, was der Vielseitigkeit des Systems abträglich ist. Insbesondere die Sakkadenanalyse wäre für Eye Tracking Studien im EC 135 Cockpit in Zukunft interessant. Neben den fehlenden Blickparametern muss außerdem angemerkt werden, dass bei der Auswertung der Blickvideos an jeder Stelle **Dateneingriffe durch den Versuchsleiter** getätigt werden können oder sogar müssen. Dies hat zwar auswertungsbezogene Vorteile, kann aber unter Umständen das Kriterium der Auswertungsobjektivität beeinträchtigen. Die Pupillenerkennung ist beispielsweise durch manuelles Festlegen der Pupillenposition im Blickvideo manipulierbar, Szenen aus Blickvideos können herausgeschnitten und die Position von AOIs händig eingezeichnet werden. Aus dem Blickwinkel wissenschaftlichen Arbeitens sollte diese subjektive Störquelle zumindest in Handbüchern der Hersteller redlich gemacht werden.

Außerdem sollte der Fakt Beachtung finden, dass das Eye Tracking System nicht Fixationen erfasst, sondern **Blicke (Gazes)**. Dabei gilt als kleinste Auswerteeinheit ein Block (Frame) von 40 ms. Was auch immer innerhalb der 40 ms passiert (sowohl Fixationen als auch Sakkaden können in einem definierten AOI innerhalb von 40 ms stattfinden) wird als Blick gewertet und verarbeitet. Die Verwendung von Blicken wird von Underwood und Everatt (1992, S. 117) wie folgt paraphrasiert: „Gaze provides an overall measure of attention while a single fixation duration gives a finer-grained estimate of moment-to-moment Processing: if one wants to know whether to carry a raincoat out when taking a walk the monthly rainfall figures may not provide the best evidence“. Da es bei der Analyse von Target Fixations um Wertebereiche im vierstelligen Millisekundenbereich geht, dürfte die Analyse von Blicken als gemessene Größe also hinreichend sein.

Nicht hinreichend befriedigend ist aus technischer Sicht aber, dass die **Güte der Markererkennung** darüber entscheidet, ob ein Blickvideo auswertbar oder komplett unbrauchbar ist. Ob z. B. ein Infrarot-Marker durch die Software des Eye Tracking Systems durchgehend erkannt wurde, stellt sich erst nach zeitintensiven Prüfungen nach der Datenerhebung heraus. In der vorliegenden Studie mussten aufgrund von unzureichender Markererkennung 21 von 66 Blickvideos **verworfen** werden – dies entspricht einem **Drittel der Datensätze** und ist als ausgesprochen kritisch zu bewerten. Die Markererkennung wird dabei nicht nur durch wechselnde Lichtbedingungen beeinflusst, sondern auch durch die Position des Markers im

Sichtfeld des Probanden und die Schärfeeinstellung der Feldkamera. Da Arbeitsplätze wie beispielsweise Cockpitsimulatoren, aber auch Kontrollstationen mit PC-Anwendung häufig schwierige Lichtbedingungen haben, ständige Kopfbewegungen der Operateure immanent aber relevant für die Analyse von Scanning Techniken sind, besteht hier Verbesserungs- bzw. Entwicklungsbedarf⁴⁴.

Eine andere Auffälligkeit bei der Datenauswertung ergab sich bei der Verarbeitung der 66 Datensätze durch die Auswertungs-Software D-Lab als 66 Blickverlaufsdigramme (Strichcodes) erstellt werden sollten. So gelang es selbst nach **48-stündiger Rechenzeit** nicht, vollständige Blickverlaufsdigramme für alle 66 Fälle zu extrahieren, da die zugehörige PDF-Datei stets **unvollständig** ausgegeben wurde. Hier ist von einem Software- bzw. Programmierfehler auszugehen oder aber von unzureichendem Arbeitsspeicher des Aufnahmerechners, welcher beim Konvertieren der Daten auf speicherintensive Auslagerungen auf die Festplatte zurückgreifen muss. Da die Blickverlaufsdigramme für die vorliegende inferenzstatistische Auswertung von Target Fixations allerdings nicht essentiell sind, war deren Erstellung hier obsolet. Der beschriebene Punkt wird aber dann virulent, wenn unmittelbare visuelle Rückmeldung an Operateure abgegeben werden soll.

5.5 Praktische Implikationen und Forschungsausblick

5.5.1 Training

Unter der Voraussetzung der **technischen Weiterentwicklung** des Eye Tracking Systems Dikablis (Version von 2009) wäre es zukünftig denkbar, Eye Tracking im **Simulator- oder auch in das Desorientierungstraining** von Piloten zu integrieren und beispielsweise im Rahmen von **Debriefings oder Crew Ressource Management Seminaren** Scanning Techniken von Piloten auszuwerten und zu diskutieren, um den Transfer in den Realflug zu optimieren. Dabei geht es nicht nur um die Ausbildung von Flugschülern (vgl. Sullivan et al., 2011), sondern auch um kontinuierliche Rückmeldung für erfahrenere Piloten. Wenn es insbesondere gelingt,

- die Markererkennung zu verbessern,
- die Kalibrierung abzukürzen,
- die Kameras an bzw. in das Visier des Helms zu integrieren,
- die grafischen Auswertungsoptionen zu optimieren und
- Eye Tracking Systeme für den Realflug zu lizenzieren,

⁴⁴ Die Firma Ergoneers GmbH bietet inzwischen Software-Upgrades an („Recording Software 2.5“: Kalibrierassistenten, Pre-Trial Markererkennung, Vorabfestlegung von AOIs usw.; Marlene Gagesch, Ergoneers GmbH, am 11.6.2012 per Email-Kommunikation).

könnten die Blickvideos von guten und schlechten Performern verglichen, Anteile von Target Fixations auf ihre Koppelung mit bestimmten Ereignissen in Missionen untersucht oder gänzlich unerfahrenen Pilotenanwärtern die eigenen Scanning Techniken vor Augen geführt werden. Fragt man Fluglehrer der Heeresfliegerwafenschule diesbezüglich, wünschen sie sich mehr zeitliche Ressourcen für detaillierte Debriefings von Flugschülern und eine Schwerpunktsetzung auf individuelle Rückmeldung. Die potenzielle Effizienz eines durch Eye Tracking unterstützten Vorgehens bei der Ausbildung wurde durch Wetzel et al. (1998, S. 8) bereits beim Training von F-16 Flugschülern nachgewiesen: „[...] *the best use of the system at point was in the creation of training tapes demonstrating effective scanning behavior for various complex scenarios.*“

5.5.2 Auswahl

Eine andere Implikation ist die Anwendung des Verfahrens bei der Auswahl von potenziellen Piloten. Das Deutsche Zentrum für Luft- und Raumfahrt e. V. (DLR) macht sich diesen Gedanken bereits praktisch zu Nutze und arbeitet an einem Eye Tracking gestützten **Auswahlalgorithmus** für Fluglotsen (Fuhrmann, 2012). Dabei werden in Einzelfallbetrachtungen Scanning Techniken (hier: Blickwege) von Probanden mit einem Modell von erfahrenen Fluglotsen abgeglichen. In Zukunft soll damit die Eignung als Fluglotse detektiert werden: „*Die Verweildauer der Blicke auf wichtigen Ereignissen hängt signifikant davon ab, ob der Bewerber den Verkehr später steuern kann*“ (ebd., S. 20).

5.5.3 Forschung

Um wissenschaftliche Aussagen darüber treffen zu können, inwiefern Eye Tracking Daten Einblicke in kognitive Prozesse liefern können, sollte für weitere Forschungsarbeiten ein **hypothesengeleitetes** Vorgehen gegenüber rein explorativen Untersuchungen favorisiert werden. Tabelle 21 schlägt hierfür einen Leitfaden vor. Resümierend gilt für alle durch Eye Tracking erhobenen Daten und Erkenntnisse, dass sie durch replizierende Arbeiten validiert und auf ihre Generalisierbarkeit – im fliegerischen Bereich insbesondere im Realflug – geprüft werden müssen.

Tabelle 21: Leitfaden zum standardisierten Vorgehen bei Eye Tracking Studien.

Methodischer Baustein	Beispielfrage
Schaffung eines theoretischen Rahmens	<ul style="list-style-type: none"> • Wodurch werden die zu messenden Blickparameter determiniert? (Endogene vs. exogene visuelle Aufmerksamkeit?)
Berücksichtigung der Grundannahmen	<ul style="list-style-type: none"> • Blick = Ort der Aufmerksamkeit oder nicht (Eye-Mind-Hypothese)? • Foveales vs. parafoveales und peripheres Sehen? • Welcher Blickparameter indiziert was (z. B. Ist ein langer Blick Zeichen für kognitiven Ressourceneinsatz oder für Langeweile)?
Definition eines experimentellen Paradigmas	<ul style="list-style-type: none"> • Experten vs. Novizen? • Hohe vs. geringe Anforderungen? • Instruktion vs. keine Instruktion der Probanden?
Ableitung von Hypothesen	<ul style="list-style-type: none"> • Partialhypothesen (Singuläreinflüsse) oder multivariates Modell?
Auswahl eines Eye Tracking Systems	<ul style="list-style-type: none"> • Remote vs. Head Mounted Eye Tracking System?
Berücksichtigung technischer Herausforderungen	<ul style="list-style-type: none"> • Aufwand vs. Ergebnis?
Datenauswertung nach klar definierten Indikatoren	<ul style="list-style-type: none"> • Theoretische oder empirische Ableitung des Indikators?
Ergebnisinterpretation im festgelegten Paradigma	<ul style="list-style-type: none"> • Kontextspezifität vs. Generalisierbarkeit?

6 Literaturverzeichnis

- Ablassmeier, M., Poitschke, T., Wallhoff, F., Bengler, K., & Rigoll, G. (2007). Eye Gaze Studies Comparing Head-Up and Head-Down Displays in Vehicles. In IEEE International Conference on Multimedia and Exposition (pp. 2250-2252). Beijing, China.
- Akinwuntan, A.E., DeWeerd, W., Feys, H., Baten, G., Arno, P., & Kiekens, C. (2003). Reliability of a road test after stroke. *Archives of Physical Medicine and Rehabilitation*, 84(12), 1792-1796.
- Alliance of Automobile Manufacturers (AAM). (2002). Driver Focus-Telematics Working Group: Statement of principles, criteria and verification procedures on driver interactions with advanced in-vehicle information and communication systems. Washington DC: Alliance of Automobile Manufacturers (AAM).
- Amelang, M. & Schmidt-Azert, L. (2006). *Psychologische Diagnostik und Intervention*. Berlin: Springer.
- Anders, G. (2001). Pilot's Attention Allocation During Approach and Landing – Eye- and Head-Tracking Research in an A330 Full Flight Simulator. In International Symposium on Aviation Psychology (ISAP). Columbus, OH, USA.
- Anderson, J.R. (1989). *Kognitive Psychologie: Eine Einführung* (2. Aufl.). Heidelberg: Spektrum-der-Wissenschaft-Verlagsgesellschaft.
- Anderson, J.R., Bothell, D., & Douglass, S. (2004). Eye Movements Do Not Reflect Retrieval Processes. *Limits of the Eye-Mind Hypothesis*. *Psychological Science*, 15(4), 225-31.
- Baddeley, A.D. & Hitch, G.J. (1974). Working memory. In G. Bower (Ed.), *Recent Advances in Learning and Motivation* (vol. 8). New York: Academic Press.
- Bak, P.M. (1999). *Selektive Aufmerksamkeit und kognitive Ressourcen. Untersuchungen zum flanker-Effekt*. Dissertation. Universität Trier, Fachbereich Psychologie.
- Balcetis, E., & Dunning, D. (2006). See what you want to see: Motivational influences on visual perception. *Journal of Personality and Social Psychology*, 91(4), 612.
- Bayram, S., & Esgin, E. (2013). Patterns of Visual Attention and Gaze to Human and Animal Faces in Children with Autism Spectrum Disorders. *World Journal on Educational Technology*, 4(3), 165-176.
- Beatty, J. (1982). Task-Evoked Pupillary Responses, Processing Load, and the Structure of Processing Resources. *Psychological Bulletin*, 91, 276-292.
- Bellenkes, A.H., Wickens, C.D., & Kramer, A.F. (1997). Visual scanning and pilot expertise: The role of attentional flexibility and mental model development. *Aviation, Space, and Environmental Medicine*, 68(7), 569-579.
- Bierhoff, H.-W. & Frey, D. (2011). *Sozialpsychologie – Individuum und soziale Welt*. Göttingen: Hogrefe.
- Billings, D.R. (2008). The Effects of Input Device and Latency on Ability to Effectively Pilot a Simulated Micro-UAV. In Proceedings of the Human Factors and Ergonomics Society 52nd Annual Meeting (pp. 2092-2096). New York, USA.
- Borman, W.C. (1977). Consistency of rating accuracy and rating errors in the judgment of human performance. *Organizational Behavior and Human Performance*, 20(2), 238-252.
- Bortz, J. (1993). *Statistik für Sozialwissenschaftler* (4. Aufl.) Berlin: Springer.
- Bradley, M.M., Keil, A., & Lang, P.J. (2012). Orienting and emotional perception: Facilitation, attenuation, and interference. *Frontiers in Psychology*, 3, 493-499.
- Brewster, S.A. & Cryer, P.G. (1999). Maximising screen-space on mobile computing devices. In Conference on Human Factors in Computing Systems (pp. 224-225). 15-20 May 1999, Pittsburgh, Pennsylvania, USA.
- Broadbent, D.E. (1958). *Perception and communication*. New York: Oxford University Press.
- Buld, S. (2000). *Der Arbeitsplatzinhaber als Experte bei der Arbeitsplatzbewertung*. Dissertation. Universität Würzburg, Würzburg.

- Byers, J.C., Bittner, A.C., & Hill, S.G. (1989). Traditional and raw task load index (TLX) correlations: Are paired comparisons necessary? In A. Mital (Ed.), *Advances in Industrial Ergonomics and Safety* (pp. 481-485). London: Taylor & Francis.
- Byrne, M.D. (2001). ACT-R/PM and Menu Selection: Applying a Cognitive Architecture to HCI. *International Journal of Human Computer Studies*, 55, 41-84.
- Cain, B. (2007). A Review of the Mental Workload Literature. Online-Dokument verfügbar unter: <http://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA474193> [Zugriff: 24.07.2012].
- CD zur Dissertation (2013). Weiterführende statistische Analysen zur Eye Tracking Studie. Flugmedizinisches Institut der Luftwaffe, Abteilung Ergonomie, Manching.
- Changizi, M.A., & Hall, W.G. (2001). Thirst modulates a perception. *Perception*, 30, 1489–1497.
- Cherry, E.C. (1953). Some experiments on the recognition of speech, with one and with two ears. *Journal of the Acoustical Society of America*, 25, 975-979.
- Chin, C., Lee, S., & Ramey, J. (2005). An Orientation to Eye Tracking in Usability Studies. *Usability and Information Design*, 273-276.
- Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. San Diego, CA, USA: Academic Press.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (vol. 2). Hillsdale, NJ: Erlbaum.
- Coladarci, T. (1986). Accuracy of teacher judgments of student responses to standardized test items. *Journal of Educational Psychology*, 78(2), 141-146.
- Colvin, K.W., Dodhia, R.M., Bechler, S.A., & Dismukes, R.K. (2003). Scanning for Visual Traffic: An Eye Tracking Study. In 12th International Symposium on Aviation Psychology (pp. 255-260). Dayton, OH, USA: The Wright State University.
- Colvin, K.W., Dodhia, R.M., & Dismukes, R.K. (2005). Is Pilots' Visual Scanning Adequate to Avoid Mid-Air Collisions? In 13th International Symposium on Aviation Psychology (pp. 104-109). Oklahoma City, USA.
- Crane, P.M. (1992). Theories of expertise as models for understanding situation awareness. In Proceedings of the 13th Annual Symposium on Psychology in the Department of Defense (pp. 148-152). Colorado Springs, Colorado, USA.
- Datta, R., & DeYoe, E. (2009). I know where you are secretly attending! The topography of human visual attention revealed with fMRI. *Vision Research*, 49, 1037–1044.
- Davidson, P.R., & Parker, K.C.H. (2001). Eye movement desensitization and reprocessing (EMDR): A meta-analysis. *Journal of Consulting and Clinical Psychology*, 69, 305–316.
- Denning, T., Bennett, W., & Crane, P. (2003). *Mission Complexity Scoring in Distributed Mission Training*. United States Air Force Research Laboratory (Rep. No. AFRL-HE-AZ-TR-2002-0192), Ohio, USA.
- DeSantis, R., Zhou, Q., & Ramey, J. (2005). A Comparison of Eye Tracking Tools in Usability Testing. *Usability and Information Design*, 210-213.
- Deutsch, J.A., & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review*, 70, 80-90.
- DeWaard, D. (1996). *The Measurement of Drivers' Mental Workload*. Haren, The Netherlands: The Traffic Research Centre VSC, University of Groningen.
- Dikablis Cable und D-Lab (2011). Eye Tracking System und Softwarepaket für Windows. Informationen verfügbar unter: <http://www.ergoneers.com/de/products/dlab-dikablis/overview.html> [Zugriff: 22.03.2012].
- Dillman, D.A. (2000). *Mail and internet surveys: The tailored design method*. New York: Wiley.
- DIN 33405 (1987). *Psychische Belastung und Beanspruchung: Allgemeines, Begriffe*. Berlin: Beuth.
- DiNocera, F., Camilli, M., & Terenzi, M. (2007). A Random Glance at the Flight Deck: Pilot's Scanning Strategies and the Real-Time Assessment of Mental Workload. *Journal of Cognitive Engineering and Decision Making*, 1, 271-285.
- Dionisio, D.P., Granholm, E., Hillix, W.A., & Perrine, W.F. (2001). Differentiation of Deception Using Pupillary Responses as an Index of Cognitive Processing. *Psychophysiology*, 38, 205-211.

- Dixon, K.W., Krueger, G.M., Rojas, V.A., & Martin, E.L. (1990). Visual Behavior in the F-15 Simulator for Air-to-Air Combat (Rep. No. AFHRL-TP-89-75). Williams Air Force Base, Arizona, USA: Air Force Human Resources Laboratory.
- Dixon, K.W., Rojas, V.A., Krueger, G.M., & Simcik, L. (1990). Eye Tracking Device for the Measurement of Flight Performance in Simulators (Rep. No. AFHRL-TP-89-18). Williams Air Force Base, Arizona, USA: Air Force Human Resources Laboratory.
- Dörner, D. (1993). Wissen, Emotionen und Handlungsregulation oder die Vernunft der Gefühle. *Zeitschrift für Psychologie*, 2, 167 - 202.
- Donath, D., Rauschert, A., & Schulte, A. (2010). Cognitive assistant system concept for multi-UAV guidance using human operator behavior models. In Proceedings of the Conference on Human Operating Unmanned Systems (HUMOUS'10). Toulouse, France.
- Duchowski, A.T. (2002). A Breadth-First Survey of Eye Tracking Applications. *Behavior Research Methods, Instruments, and Computing*, 34, 455-470.
- Duchowski, A.T., Medlin, E., Gramopadhye, A., Melloy, B., & Nair, S. (2001). Binocular eye tracking in VR for visual inspection training. In Proceedings of the ACM symposium on Virtual reality software and technology (pp. 1-8). ACM. Online-Dokument verfügbar unter: <http://andrewd.ces.clemson.edu/research/vislab/docs/vrst.pdf> [Zugriff: 04.02.2013].
- Ehrlinger, J. & Dunning, D. (2003). How chronic self-views influence (and potentially mislead) estimates of performance. *Journal of Personality and Social Psychology*, 84(1), 5-17.
- Endsley, M.R. (1988). Design and evaluation for situation awareness enhancement. In Proceedings of the Human Factors Society 32nd Annual Meeting (pp. 97-101). Santa Monica, CA: Human Factors and Ergonomics Society.
- Endsley, M.R. (1993). Situation awareness and workload: Flip sides of the same coin. In R.S. Jensen & D. Neumeister (Eds.), Proceedings of the 7th International Symposium on Aviation Psychology (pp. 906-911). Columbus, OH: Department of Aviation, Ohio State University.
- Endsley, M.R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors*, 37, 32-64.
- Endsley, M.R. (1996). Automation and situation awareness. In R.Parasuraman & M. Mouloua (Eds.), *Automation and Human Performance: Theory and Applications* (pp. 163-181). Mahwah, NJ, USA: Lawrence Erlbaum.
- Endsley, M.R. (2000). Theoretical underpinnings of situation awareness: A critical review. In M.R.Endsley & D.J. Garland (Eds.), *Situation Awareness Analysis and Measurement* (pp. 3-32). Mahwah, NJ, USA: Lawrence Erlbaum.
- Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments, and Computing*, 28, 1-11.
- Ergoneers GmbH. (2009). *Handbuch Dikablis Eye-Tracking System*. Ergoneers GmbH, Manching.
- Ericsson, K.A., & Charness, N. (1997). Cognitive and developmental factors in expert performance. In P.J. Feltovich, K.M. Ford, & R.R. Holman (Eds.), *Expertise in context. Human and machine* (pp. 3-41). Menlo Park, CA: AAAI Press/The MIT Press.
- Ericsson, K.A., & Delaney, P.F. (1998). Working memory and expert performance. In R.H. Logie & K.J. Gilhooly (Eds.), *Working memory and thinking* (pp. 93-114). Hove, UK: Psychology Press.
- Ericsson, K.A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, 102, 211-245.
- European Aviation Safety Agency (EASA) (2010). Collision Avoidance. Methods to reduce the risk. Onlinedokument verfügbar unter: http://easa.europa.eu/essi/documents/EGAST_Leaflet1.pdf [Zugriff: 22.03.2012].
- Fairclough, S.H. (1991). Adapting the TLX to Measure Driver Mental Workload (DRIVE Project V1017 [BERTIE] Rep. No. 71). Loughborough, UK: HUSAT Research Institute.
- Federal Aviation Administration (FAA) (1996). Human Factors Team Report on: The Interfaces Between Flightcrews and Modern Flight Deck Systems. Oklahoma City, OK, USA: Federal Aviation Administration.

- Federal Aviation Administration (FAA) (1998). Scanning for Other Aircraft. Aeronautical Information Manual (Rep. No. 8-1-6-c). Oklahoma City, OK USA: Federal Aviation Administration.
- Findlay, J.M. (2005). Covert Attention and Saccadic Eye Movements. In I. Laurent, R. Geraint, & K.T. John (Eds.), *Neurobiology of Attention* (pp. 114-116). Burlington: Academic Press.
- Findlay, J.M., & Gilchrist, I.D. (2003). *Active Vision. The Psychology of Looking and Seeing*. New York: Oxford.
- Flemisch, O.F. (2001). *Pointillistische Analyse der visuellen und nicht-visuellen Interaktionsressourcen am Beispiel Pilot-Assistenzsystem*. Dissertation. Universität der Bundeswehr München, Aachen: Shaker Verlag.
- Flugmedizinisches Institut der Luftwaffe (2012). Abteilung Ergonomie. Manching, Deutschland.
- Flugmedizinisches Institut der Luftwaffe; Pongratz, H. (Hrsg.) (2006). *Kompendium der Flugmedizin*. Flugmedizinisches Institut der Luftwaffe, Fürstenfeldbruck.
- Fox, J., Merwin, D., Marsh, R., McConkie, G., Kramer, A. (1996). Information extraction during instrument flight: An evaluation of the validity of the eye-mind hypothesis. *Proceedings of the Human Factors and Ergonomics Society 40th Annual Meeting*, Philadelphia, Pennsylvania, USA.
- Fuhrmann, L. (2012). Tests mit Perspektive. In Hamburg entwickeln DLR-Wissenschaftler Messverfahren für die Eignung der Fluglotsen und Piloten von morgen. *Magazin des Deutschen Zentrums für Luft- und Raumfahrt e.V. (DLR)*, 135, 19-21.
- Geise, S. (2011). *Vision that matters. Die Wirkungs- und Funktionslogik Visueller Politischer Kommunikation am Beispiel des Wahlplakats*. Wiesbaden: VS.
- Geise, S. (2012). *Eyetracking in der Kommunikations- und Medienwissenschaft: Theorie, Methode und kritische Reflexion*. Online-Dokument verfügbar unter <http://www.scm.nomos.de/aktuelles-heft-und-archiv/2011/heft-2/beitrag-geise/> [Zugriff: 19.06.2012].
- Gibson, J., Orasanu, J., Villeda, E., & Nygren, T. (1997). Loss of situation awareness- Causes and consequences. *Proceedings of the 9th International Symposium on Aviation Psychology (ISAP)*. Columbus, OH, USA.
- Glucksberg, S. & Cowen, J.G.N. (1970). Memory for Nonattended Auditory Material. *Cognitive Psychology*, 1, 149–156.
- Goldberg, T.E., Berman, K.F., Fleming, K., Ostrem, J., Van Horn, J.D., Esposito, G., Mattay, V.S., Gold, J.M., & Weinberger, D.R. (1998). Uncoupling cognitive workload and prefrontal cortical physiology: A PET rCBF study. *Neuroimage*, 7, 296-303.
- Gopher, D. & Donchin, E. (1986). Workload – An examination of the concept. In K.R. Boff, L. Kaufman, & J.P. Thomas (Eds.), *Handbook of Human Perception and Performance, Vol. 2, Cognitive Processes and Performance* (pp. 41-1–41-49). New York: Wiley.
- Graeber, D.A. & Andre, A.D. (1999). Assessing Visual Attention of Pilots while Using Electronic Moving Maps for Taxiing. In S. Jensen, B. Cox, J. D. Callister, & R. Lavis (Eds.), *10th Symposium on Aviation Psychology* (pp. 791-796). Columbus, Ohio, USA: Ohio State University.
- Grandt, M. (2004). *Erfassung und Bewertung der mentalen Beanspruchung mittels psychophysiologischer Messverfahren (FKIE-Bericht Nr. 88)*. Forschungsgesellschaft für Angewandte Naturwissenschaften e. V. (FGAN), Forschungsinstitut für Kommunikation, Informationsverarbeitung und Ergonomie, Wachtberg.
- Grey, J. & Wedderburn, A. (1960). Grouping Strategies with Simultaneous Stimuli. *Quarterly Journal of Experimental Psychology*, 12, 180-195.
- Grier, R. (2008). The redline of workload: Theory, research and design. In *52nd Annual Meeting of the Human Factors and Ergonomics Society* (pp. 1204-1208). New York, USA: Human Factors & Ergonomics Society.
- Grube, H. (2007). *Ausbildungshandbuch für die Hubschrauberführergrundausbildung Teil VFR. Kommandeur der Heeresfliegerwaffenschule, Lehrgruppe A. Schäferkaserne, Bückeburg*.
- Hacker, W. (1973). *Allgemeine Arbeits- und Ingenieurpsychologie*. VEB Deutscher Verlag der Wissenschaften.

- Hart, S.G. (2006). NASA-Task Load Index (NASA-TLX); 20 years later. In Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting (pp. 904-908). Santa Monica, CA: Human Factors & Ergonomics Society.
- Hart, S.G. & Staveland, L.E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In P. A.Hancock & N. Meshkati (Eds.), Human Mental Workload (pp. 139-183). Amsterdam: North-Holland.
- Hart, S.G., & Wickens, C.D. (1990). Workload assessment and prediction. In H.R. Booher (Ed.), MAN-PRINT: An approach to systems integration (pp. 257-296). New York: Van Nostrand Reinhold.
- Hayashi, M. (2003). Hidden Markov Models to Identify Pilot Instrument Scanning and Attention Patterns. In IEEE International Conference on Systems, Man, & Cybernetics (pp. 2889-2896). Washington D.C., USA.
- Hayashi, M., Oman, C.M., & Zuschlag, M. (2004). Effects of Head-Up Display Airspeed Indicator and Altimeter Formats on Pilot Scanning and Attention Switching. In 5th Conference on Human Performance, Situation Awareness, and Automation Technology. Daytona Beach, Florida, USA.
- Hayhoe, M. (2004). Advances in Relating Eye Movements and Cognition. *Infancy*, 6(2), 267-274.
- Hayhoe, M. & Ballard, D. (2005). Eye movements in natural behavior. *Trends in Cognitive Sciences*, 9(4), 188-194.
- Hayhoe, M., Shrivastava, A., Myruczek, R., & Pelz, J. (2003). Visual Memory and Motor Planning in a Natural Task. *Journal of Vision*, 3, 49-63.
- Hendee, W.R. & Wells, P.N.T. (1993). *The Perception of Visual Information*. New York: Springer.
- Henderson, J.M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences*, 7(11), 498-504.
- Hennessey, C. & Lawrence, P. (2008). Improving the Accuracy and Reliability of Remote System-Calibration-Free Eye-gaze Tracking. Proceeding of IEEE Transactions on Biomedical Engineering. Online-Dokument verfügbar unter: <http://hcggroup.org/images/stories/Files/Publications/2009/Hennessey2009b.pdf> [Zugriff: 31.01.2013].
- Henning, M. (2010). Preparation for Lane Change Manoeuvres: Behavioral Indicators and Underlying Cognitive Processes. Dissertation. TU Chemnitz, Fakultät für Human- und Sozialwissenschaften, Chemnitz.
- Herczeg, M. (2005). *Software-Ergonomie: Theorien, Modelle und Kriterien für gebrauchstaugliche interaktive Computersysteme* (3. Aufl.) München: Oldenbourg Wissenschaftsverlag.
- Herczeg, M. & Stein, M. (2012). Human Aspects of Information Ergonomics. In M. Stein & P. Sandl (Eds.). *Information Ergonomics: A theoretical approach and practical experience in transportation* (pp. 59-98). Berlin: Springer.
- Hoffman, J.E. (1999). Stages of Processing in Visual Search and Attention. In B.H. Challis & B.M. Velichkovsky (Eds.), *Stratification in Cognition and Consciousness* (pp. 43-71). Amsterdam: John Benjamins Publishing.
- Horberr, T., Anderson, J., Regan, M.A., Triggs, T.J., & Brown, J. (2006). Driver distraction: The effects of concurrent in-vehicle tasks, road environment complexity and age on driving performance. *Accident Analysis and Prevention*, 38, 185-191.
- Horend, S. (2004). Integration eines Blickbewegungsmesssystems in eine einfache Flugsimulationsumgebung. Nicht veröffentlichte Projektarbeit. EADS Deutschland GmbH und Berufsakademie Ravensburg.
- Horrey, W.J., Wickens, C.D., & Consalus, K.P. (2006). Modeling drivers' visual attention allocation while interacting with in-vehicle technologies. *Journal of Experimental Psychology: Applied*, 12, 67-86.
- Hudlicka, E., & McNeese, M.D. (2002). Assessment of user affective and belief states for interface adaptation: Application to an Air Force pilot task. *User Modeling and User-Adapted Interaction*, 12(1), 1-47.
- Hughes, E.R., Hassoun, J.A., Ward, G.F., & Rueb, J.D. (1990). An assessment of selected workload and situation awareness metrics in a part-mission simulation: Final report (ASD-TR-90-5009). Wright-Patterson AFB, OH: Aeronautical Systems Division, Air Force Systems Command.

- Inhoff, A.W., Briehl, D., & Schwartz, J. (1996). Compound word effects differ in reading, on-line naming, and delayed naming tasks. *Memory & Cognition*, 24, 466-476.
- Inhoff, A.W. & Radach, R. (1998). Definition and computation of oculomotor measures in the study of cognitive processes. In G. Underwood (Ed.), *Eye Guidance in Reading and Scene Perception* (pp. 29-53). Oxford: Elsevier.
- ISO 10075-1 (1991). Ergonomic principles related to mental workload. Part 1: General terms and definitions.
- ISO 15007-1 (2002). Road vehicles – Measurement of driver visual behaviour with respect to transport information and control systems. Part 1: Definitions and parameters.
- ISO 15007-2 (2001). Road vehicles – Measurement of driver visual behaviour with respect to transport information and control systems. Part 2: Equipment and procedures.
- ISO 16673 (2006). Road vehicles – Ergonomic aspects of transport information and control systems – Occlusion method to assess visual demand due to the use of in-vehicle systems.
- Jacob, R. & Karn, K. (2003). Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*, 573-605.
- Jako, R.A., & Murphy, K.R. (1990). Distributional ratings, judgment decomposition, and their impact on interrater agreement and rating accuracy. *Journal of Applied Psychology*, 75(5), 500.
- Janelle, C.M. & Hatfield, B.D. (2008). Visual attention and brain processes that underlie expert performance: Implications for sport and military psychology. *Military Psychology*, 20, S39-S69.
- Janssen, J. & Laatz, W. (2010). *Statistische Datenanalyse mit SPSS. Eine anwendungsorientierte Einführung in das Basissystem und das Modul Exakte Tests* (7. Aufl.). Berlin: Springer.
- Jönsson, E. (2005). *If Looks Could Kill - An Evaluation of Eye Tracking in Computer Games*. Master Thesis. Stockholm Royal Institute of Technology, Department of Numerical Analysis and Computer Science, Stockholm, Sweden.
- Johannsen, G. (1979). Workload and Workload Measurement. In N. Moray (Ed.), *Mental Workload, Its Theory and Measurement* (pp. 3-11). New York: Plenum Press.
- Johannsen, G. (1993). *Mensch-Maschine-Systeme*. Berlin: Springer.
- Joiko, K., Schmauder, M., & Wolff, G. (2012). Psychische Belastung und Beanspruchung im Berufsleben: Erkennen – Gestalten. Online-Dokument verfügbar unter: <http://www.baua.de/cae/ servlet/contentblob/673898/publicationFile/50008/A45.pdf> [Zugriff: 26.05.2012].
- Jones, D.G. & Endsley, M.R. (1996). Sources of situation awareness errors in aviation. *Aviation, Space, and Environmental Medicine*, 67(6), 507-512.
- Joyce, C.A., Gorodnitsky, I.F., & Kutas, M. (2004). Automatic removal of eye movement and blink artifacts from EEG data using blind component separation. *Psychophysiology*, 41(2), 313-325.
- Just, M.A. & Carpenter, P.A. (1980). A Theory of Reading: From Eye Fixations to Comprehension. *Psychological Review*, 87, 329-354.
- Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ, USA: Prentice-Hall.
- Kanwisher, N. (1987). Repetition blindness: Type recognition without token individuation. *Cognition*, 27, 117-143.
- Kasarskis, P., Stehwien, J., Hickox, J., Aretz, A., & Wickens, C.D. (2001). Comparison of expert and novice scan behaviors during VFR flight. *Proceedings of the 11th International Symposium on Aviation Psychology*. Columbus (OH): The Ohio State University.
- Keller, P.M. (2004). *Untersuchung der Pupillenreaktion und -oszillation an einem Normkollektiv mittels Compact Integrated Pupillograph (CIP) der Firma AMTech*. Dissertation. Ruhr-Universität Bochum, Medizinische Fakultät, Bochum.
- Kennedy, R.S., Lane, N.E., Berbaum, K.S., & Lillenthal, M.G. (1993). A simulator sickness questionnaire (SSQ): A new method for quantifying simulator sickness. *International Journal of Aviation Psychology*, 3(3), 203-220.
- Kerr, B. (1973). Processing demands during mental operations. *Memory & Cognition*, 1, 401-412.

- Kimble, M.O., Fleming, K., Bandy, C., Kim, J., & Zambetti, A. (2010). Eye tracking and visual attention to threatening stimuli in veterans of the Iraq war. *Journal of Anxiety Disorders*, 24(3), 293-299.
- Klein, M. & Arzheimer, K. (2012). Ranking- und Rating-Verfahren zur Messung von Wertorientierungen, untersucht am Beispiel des Ingelhart-Index. Empirische Befunde eines Methodenexperimentes. Online-Dokument verfügbar unter: <http://www.kai-arzheimer.com/ingelhart-panel/ingelhart-panel.html> [Zugriff: 27.05.2012].
- Klingner, J., Kumar, R., & Hanrahan, P. (2008). Measuring Task-Evoked Pupillary Response with a Remote Eye Tracker. In 2008 Symposium on Eye Tracking Research & Applications (pp. 69-72). Savannah, Georgia, USA: ACM.
- LaBerge, D. (1983). Spatial extend of attention to letters in words. *Journal of Experimental Psychology: Human Perception & Performance*, 9, 371-379.
- Land, M.F., & Furneaux, S. (1997). The Knowledge Base of the Oculomotor System. *Philosophical Transactions of the Royal Society London. Biological Sciences*, 382, 1231-1239.
- Land, M.F., & McLeod, P. (2000). From Eye Movements to Actions: How Batsmen Hit the Ball. *Nature America*, 3, 1340-1345.
- Land, M., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception*, 28(11), 1311-1328.
- Land, M.F., & Lee, D.N. (1994). Where we look when we steer. *Nature*, 369(6483), 742-744.
- Lee, Y.H., & Liu, B.S. (2003). Inflight workload assessment: Comparison of subjective and physiological measurements. *Aviation, Space, and Environmental Medicine*, 74(10), 1078-1084.
- Lin, T., Imamiya, A., & Mao, X. (2008). Using multiple data sources to get closer insights into user cost and task performance. *Interacting with Computers*, 20(3), 364-374.
- Lindo, R.S., Deaton, J.E., Cain, J.H., & Lang, C. (2012). Methods of Instrument Training and Effects on Pilots' Performance with Different Types of Flight Instrument Displays. *Aviation Psychology and Applied Human Factors*, 2(2), 62-71.
- Liu, L., Watson, B., & Miyazaki, M. (1999). VR for the elderly: Quantitative and qualitative differences in performance with a driving simulator. *Cyberpsychology and Behavior*, 2, 567-576.
- Mabe, P. A., & West, S. G. (1982). Validity of self-evaluation of ability: A review and meta-analysis. *Journal of Applied Psychology*, 67(3), 280-296.
- Matessa, M. & Remington, R. (2005). Eye Movements in Human Performance Modeling of Space Shuttle Operations. In *Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting* (pp. 1114-1118). Orlando, Florida, USA.
- Merchant, S. (2001). *Eye Movement Research in Aviation and Commercially Available Eye Trackers Today* (Report: Eye Movement Summary - Assessing Human Visual Performance). Iowa, USA: University of Iowa, Department of Industrial Engineering.
- Metz, B. (2009). *Worauf achtet der Fahrer? Steuerung der Aufmerksamkeit beim Fahren mit visuellen Nebenaufgaben*. Dissertation. Julius-Maximilian-Universität Würzburg, Philosophische Fakultät.
- Metzger, U. & Parasuraman, R. (2005). Automation in Future Air Traffic Management: Effects of Decision Aid Reliability on Controller Performance and Mental Workload. *Human Factors*, 47, 1-15.
- Miyake, S. (2001). Multivariate workload evaluation combining physiological and subjective measures. *International Journal of Psychophysiology*, 40(3), 233-238.
- Microsoft Excel (2003). CD-Rom für Windows XP. Microsoft Software.
- Monk, C.A. & Kidd, D.G. (2007). R We Fooling Ourselves: Does the Occlusion Technique Shortchange R Estimates? In 4th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design (pp. 2-8). Iowa City, IA, USA: University of Iowa, Public Policy Center.
- Moorthy, K., Munz, Y., Adams, S., Pandey, V., & Darzi, A. (2006). Self-assessment of performance among surgical trainees during simulated procedures in a simulated operating theater. *The American Journal of Surgery*, 192(1), 114-118.

- Morgenstern, C. (1896). Stufen Natur. Online-Quelle verfügbar unter: <http://gutzitiert.de> [Zugriff: 12.06.2012].
- Moroney, W.F., Biers, D.W., & Eggemeier, F.T. (1995). Some measurement and methodological considerations in the application of subjective workload measurement techniques. *International Journal of Aviation Psychology*, 5, 87-106.
- Mourant, R.R. & Rockwell, T.H. (1972). Strategies of visual search by novice and experienced drivers. *Human Factors*, 14(4), 325-335.
- Navon, D. & Gopher, D. (1979). On the economy of the human processing system. *Psychological Review*, 86, 214-255.
- Neisser, U. (1967). *Cognitive Psychology*. Englewood Cliffs, New Jersey: Prentice-Hall.
- Nelson, W.T., Hettinger, L.J., Cunningham, J.A., Brickman, B.J., Haas, M.W., & McKinley, R.L. (1998). Effects of Localize Auditory Information on Visual Target Detection Performance Using A Helmet-Mounted Display (Rep. No. AFRL-HE-WP-TR-1998-0117). United States Air Force Research Laboratory, Human Effectiveness Directorate, Crew System Interface Division.
- Neumann, O. (1996). Theories of attention. In O. Neumann & A.F. Sanders (Eds.), *Handbook of Perception and Action*, Vol. 3: Attention (pp. 389-446). London: Academic Press.
- Nilsson, S., Gustafsson, T., & Carleberg, P. (2009). Hands Free Interaction with Virtual Information in a Real Environment: Eye Gaze as an Interaction Tool in an Augmented Reality System. *PsychNology Journal*, 7, 175-196.
- Noble, A. (2012). Schweizer Käse mit Eis (Deutsche Übersetzung). *Das Magazin des Bundespolizei-Flugdienstes „Flugsicherheit“*, 1, 40 – 41.
- Norman, D.A., & Bobrow, D.G. (1975). On data-limited and resource limited processes. *Cognitive Psychology*, 7, 44-64.
- Nunnally, J.C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.
- O'Donnell, C.R., & Eggemeier, F.T. (1986). Workload assessment methodology. In K.R. Boff, L. Kaufman, & J.P. Thomas (Eds.), *Handbook of perception and human performance*, Vol. 2, Cognitive processes and performance (pp. 42.1-42.29). New York: Wiley.
- O'Hare, D. (2002). Aeronautical decision making: Metaphors, models, and methods. In P.S.Tsang & M.A. Vidulich (Eds.), *Principles and practices of aviation psychology* (pp. 201-237). Mahwah, NJ: Lawrence Erlbaum.
- Pannasch, S., Helmert, J.R., Roth, K., Herbold, A.K., & Walter, H. (2008). Visual fixation durations and saccade amplitudes: Shifting relationship in a variety of conditions. *Journal of Eye Movement Research*, 2(2): 4,1-19.
- Pfeiffer, T., Donner, M., Latoschik, M. E., & Wachsmuth, I. (2007). Blickfixationstiefe in stereoskopischen VR-Umgebungen: Eine vergleichende Studie. Online-Dokument verfügbar unter: <http://www.techfak.uni-bielefeld.de/~tpfeiffe/pubs/2007 - Pfeiffer Donner Latoschik Wachsmuth Blickfixationstiefe.pdf> [Zugriff: 04.02.2013].
- Pinel, J.P. (2007). *Biopsychologie* (Übers. v. P. Pauli) (6. Aufl.). München: Pearson Studium.
- Poitschke, T.M. (2011). Blickbasierte Mensch-Maschine Interaktion im Automobil. Dissertation. Technische Universität München, Lehrstuhl für Mensch-Maschine Kommunikation. Online-Dokument verfügbar unter: <http://www.mmk.ei.tum.de/publ/pdf/diss/poitschke.pdf> [Zugriff: 31.01.2013].
- Poole, A. & Ball, L.J. (2000). Eye Tracking in Human Computer Interaction and Usability Research: Current Status and Future Prospects. In C. Ghaoui (Ed.), *Encyclopedia of Human Computer Interaction*. Pennsylvania, USA: Idea Group.
- Popescu, M.M., Romero, M., & Usart, M. (2012). Using Serious Games in adult education Serious Business for Serious People-the MetaVals game case study. Online-Dokument verfügbar unter: http://www.icvl.eu/2012/disc/icvl/documente/pdf/met/ICVL_ModelsAndMethodologies_paper12.pdf [Zugriff: 04.04.2013].
- Posner, M.I. (1978). *Chronometric explorations of mind*. Hillsdale, NJ: Lawrence Erlbaum.

- Posner, M.I. (1980). Orienting of attention. *The Quarterly Journal of Experimental Psychology*, 32, 3-25.
- Posner, M.I., Petersen, S.E., Fox, P.T., & Raichle, M.E. (1988). Localization of cognitive functions in the human brain. *Science*, 240, 1627-1631.
- Previc, F.H., Lopez, N., Ercoline, W.R., Daluz, C.M., Workman, A.J., Evans, R.H., & Dillon, N.A. (2009). The effects of sleep deprivation on flight performance, instrument scanning, and physiological arousal in pilots. *The International Journal of Aviation Psychology*, 19(4), 326-346.
- Priese, L., Schmitt, F., & Lemke, P. (2007). Automatische See-Through Kalibrierung. Arbeitsbericht Nr. 7/2007 des Fachbereichs Informatik, Institut für Computervisualistik, Universität Koblenz-Landau. Online-Dokument verfügbar unter: http://www.uni-koblenz.de/~fb4reports/2007/2007_07_Arbeitsberichte.pdf [Zugriff: 01.02.2013].
- Rasmussen, J. (1986). *Information Processing and Human-Machine Interaction. An Approach to cognitive Engineering*. New York: North-Holland.
- Rayner, K. (1985). Eye Movements in Reading and Information Processing: 20 Years of Research. *Psychological Bulletin*, 124, 372-422.
- Regen, F. (2009). Assoziation zwischen Pupillen-Unruhe-Index (PUI) und Korrelaten des zentralnervösen Aktivierungsniveaus im Wach-EEG. Dissertation. Freie Universität Berlin, Medizinische Fakultät, Berlin.
- Reid, G.B. & Nygren, T.E. (1988). The subjective workload assessment technique: A scaling procedure for measuring mental workload. In P.A.M. Hancock & N. Amsterdam (Eds.), *Human Mental Workload* (pp. 185-218). North-Holland: Elsevier Science Publishers BV.
- Rohani, S.S. (2012). *Acquiring Fear and Threat Related Attentional Biases Through Informational Learning*. Dissertation. University of Sussex.
- Rohmert, W. (1984): Das Belastungs- Beanspruchungskonzept: *Zeitschrift für Arbeitswissenschaft*, 38, S. 193-200.
- Rohmert, W. & Rutenfranz, J. (1975). *Arbeitswissenschaftliche Beurteilung der Belastung und Beanspruchung aus unterschiedlichen industriellen Arbeitsplätzen*. Bonn: Bundesministerium für Arbeit und Sozialordnung.
- Roscoe, A.H. (1992). Assessing pilot workload: Why measure heart rate, HRV, and respiration? *Biological Psychology*, 34, 259-288.
- RTO/NATO (2001). *NATO Guidelines on Human Engineering Testing and Evaluation*. Online-Dokument verfügbar unter: [http://ftp.rta.nato.int/public//PubFullText/RTO/TR/RTO-TR-021//TR-021-\\$\\$TOC.pdf](http://ftp.rta.nato.int/public//PubFullText/RTO/TR/RTO-TR-021//TR-021-$$TOC.pdf) [Zugriff: 22.03.2012].
- Rubio, S., Diaz, E., Martin, J., & Puente, J.M. (2004). Evaluation of subjective mental workload: A comparison of SWAT, NASA-TLX, and workload profile methods. *Applied Psychology: An International Review*, 53, 61-86.
- Salas, E., Wilson, K.A, Burke, C.S., Wightman, D.C., & Howse, W.R. (2006). Crew resource management training research, practice, and lessons learned. In R.C. Williges (ed.), *Review of human factors and ergonomics* (pp. 35-73). Santa Monica, CA: Human Factors and Ergonomics Society.
- Salvucci, D.D. (2000). A Model of Eye Movements and Visual Attention. In 3rd International Conference on Cognitive Modeling (pp. 252-259). Veenendaal, The Netherlands: Universal Press.
- Sanders, A.F. (1983). Towards a model of stress and human performance. *Acta Psychologica*, 53(1), 61-97.
- Sarter, N.B. & Woods, D.D. (1991). Situation Awareness: A Critical But Ill Defined Phenomenon. *International Journal of Aviation Psychology*, 1, 45-57.
- Schäfer, C. (2007). Demonstrator für Interaktion durch Augenbewegung II. Studienarbeit. Universität Koblenz-Landau, Fachbereich Informatik.
- Schandry, R. (1989). *Lehrbuch Psychophysiologie: Körperliche Indikatoren psychischen Geschehens* (2. Aufl.). Weinheim: Psychologie Verlags Union.

- Schönpflug, W. (1987): Beanspruchung und Belastung bei der Arbeit – Konzepte und Theorien. In U. Kleinbeck & J. Rutenfranz (Hrsg.), *Enzyklopädie der Psychologie*. Themenbereich D, Serie III, Bd. 1 Arbeitspsychologie (S. 130-184). Göttingen: Hogrefe.
- Schuler, H. (2007). *Lehrbuch Organisationspsychologie* (4., aktualisierte Aufl.). Bern: Huber.
- Schultheis, H. (2004). Die Pupillengröße als Index zur Online-Erfassung der kognitiven Belastung. Online-Dokument verfügbar unter: http://cindy.informatik.uni-bremen.de/cosy/staff/schultheis/pub/Diplomarbeit_Psychologie.pdf [Zugriff: 17.10.2012].
- Schweigert, M. (2002). Fahrerblickverhalten und Nebenaufgaben. Dissertation. Technische Universität München, München.
- Seitz, D.P. & Häcker, H.O. (2008). Qualitative Befundanalyse des Konstruktes Situational Awareness – A qualitative Analysis. In Bundesministerium der Verteidigung PSZ III 6 (Hrsg.), *Untersuchungen des Psychologischen Dienstes der Bundeswehr* (Jahrgang 43). Bonn: Bundesministerium der Verteidigung (BMVg).
- Shapiro, K.L., & Raymond, J.E. (1989). Training of efficient oculomotor strategies enhances skill acquisition. *Acta Psychologica*, 71, 217–242.
- Shaw, T.A., & Porter, M.A. (2012). Emotion Recognition and Visual-Scan Paths in Fragile X Syndrome. *Journal of Autism and Developmental Disorders*, 1-21.
- Spivey, M. Richardson, D. & Dale, R. (2008). The Movement of Eye and Hand as a Window into Language and Cognition. Online-Dokument verfügbar unter: http://www.evethink.org/resources/lab_papers/Spivey2008_The_Movement_of_Ey.pdf [Zugriff: 05.04.2013].
- Sprenger, R. (2008). Empirische Forschungsmethoden für die Evaluation visueller Fahrerassistenzsysteme im Kraftfahrzeug. Dissertation. Universität Paderborn, Fakultät für Kulturwissenschaften.
- SPSS 17.0 for Windows (2008). CD-ROM für Windows XP. Prentice Hall International.
- Stahl, F. & Kleinkes, M. (2005). Eye-Tracking as a method for determining the quality of headlamp light distributions? ISAL 6th International Symposium on Automotive Lighting (pp. 912-920). München: Utz.
- Stein, M. & Robinski, M. (2011a). Analyse der Auftretenshäufigkeit und Schwere von Simulator Sickness beim EC 135 Simulator. Unveröffentlichter Abschlussbericht. Manching: Flugmedizinisches Institut der Luftwaffe, Abteilung IV Ergonomie.
- Stein, M. & Robinski, M. (2011b). Analyse der Auftretenshäufigkeit und Schwere von Simulator Sickness beim Flug- und Taktsimulator TORNADO. Unveröffentlichter Abschlussbericht. Manching: Flugmedizinisches Institut der Luftwaffe, Abteilung IV Ergonomie.
- Stein, M. & Robinski, M. (2011c). Analyse der Auftretenshäufigkeit und Schwere von Simulator Sickness beim P3-C-Orion Simulator. Unveröffentlichter Abschlussbericht. Manching: Flugmedizinisches Institut der Luftwaffe, Abteilung IV Ergonomie.
- Stein, M. & Robinski, M. (2012). Simulator Sickness in Flight Simulators of the German Armed Forces. *Aviation Psychology and Applied Human Factors*, 2(1), 11-19.
- Strasburger, H., Rentschler, I. & Jüttner, M. (2011). Peripheral vision and pattern recognition: A review. *Journal of Vision*, 11(5: 13), 1-82.
- Stroebe, W., Jonas, K., & Hewstone, M. R. (2002). *Sozialpsychologie: Eine Einführung*. Heidelberg: Springer.
- Sullivan, J., Yang, H. J., Day, M., & Kennedy, Q. (2011). Training Simulation for Helicopter Navigation by Characterizing Visual Scan Patterns. *Aviation, Space, and Environmental Medicine*, 82, 871-878.
- Svensson, E., Angelborg-Thanderz, M., Sjöberg, L., & Olsson, S. (1997). Information Complexity – Mental Workload and Performance in Combat Aircraft. *Ergonomics*, 40, 363-380.
- Svensson, E., Angelborg-Thanderz, M., & Wilson, G.F. (1999). Models of pilot performance for systems and mission evaluation – psychological and psychophysiological aspects. United States Air Force Research Laboratory (AFRL-HE-WP-TR-1999-0215). Wright-Patterson AFB, OH: Aeronautical Systems Division, Air Force Systems Command.

- Theeuwes, J. (1993). Visual Selective Attention. A Theoretical Analysis. *Acta Psychologica*, 83(2), 93-154.
- Theofanou, D. (2002). Maße zur Erfassung von visueller Ablenkung bei verschiedenen komplexen Streckenabschnitten. Dissertation. Universität Regensburg, Philosophische Fakultät II (Psychologie und Pädagogik).
- Thomas, L.C. & Wickens, C.D. (2004). Eye Tracking and Individual Differences in Off-Normal Event Detection when Flying with a Synthetic Vision System Display. In *Human Factors and Ergonomics Society 48th Annual Meeting*. Santa Monica, USA.
- Treisman, A.M. (1960). Contextual cues in selective listening. *Quarterly Journal of Experimental Psychology*, 12, 242-248.
- Treisman, A. & Geffen, G. (1967). Selective Attention: Perception or Response? *The Quarterly Journal of Experimental Psychology*, 19, 1-17.
- Treisman, A.M. & Gelade, G. (1980). A Feature-Integration Theory of Attention. *Cognitive Psychology*, 12, 97-136.
- Tsai, Y., Viirre, E., Strychacz, C., Chase, B., & Jung, T.-P. (2007). Task Performance and Eye Activity: Predicting Behavior Relating to Cognitive Workload. *Aviation, Space, and Environmental Medicine: Operational Applications of Cognitive Performance Enhancement Technologies*, 78, B176-B185.
- Tsang, P.S. (2002). Assessing Cognitive Aging in Piloting. In P.S. Tsang & M.A. Vidulich (Eds.), *Principles and practice of aviation psychology* (pp. 507-529). Mahwah, NJ: Lawrence Erlbaum.
- Tsang, P.S., & Velazquez, V.L. (1996). Diagnosticity and multidimensional subjective workload ratings. *Ergonomics*, 39(3), 358-381.
- Uhlarik, J. & Comerford, D.A. (2002). A Review of Situation Awareness Literature Relevant to Pilot Surveillance Functions. FAA Office of Aerospace Medicine (Rep. No. DOT/FAA/AM-02/3). U.S. Department of Transport.
- Underwood, G. & Everatt, J. (1992). The role of eye movements in reading: Some limitations of the eye-mind-assumption. In E. Chekaluk & K.R. Llewellyn (Eds.), *The role of eye movements in perceptual processes*. Amsterdam: Elsevier.
- Velichkovsky, B.M., Rothert, A., Moniotas, D., Dornhöfer, S.M., Joos, M., & Pannasch, S. (2003). Visual Fixations as a Rapid Indicator of Hazard Perception. In G.R.J. Hockey, A.W.K. Gaillard, & O. Burov (Eds.), *Operator Functional State and Impaired Performance in Complex Work Environments* (pp. 313-321). Amsterdam: IOS Press (NATO Science Series).
- Velichkovsky, B.M., Sprenger, A. & Pomplun, M. (1997). Auf dem Weg zur Blickmaus: Die Beeinflussung der Fixationsdauer durch kognitive und kommunikative Aufgaben. In R. Liskowsky, B.M. Velichkovsky, & W. Wüschmann (Hrsg.), *Software-Ergonomie*. Stuttgart: Teubner.
- Vidulich, M.A. (2002). Mental Workload and Situation Awareness: Essential Concepts for Aviation Psychology Practice. In P.S. Tsang & M.A. Vidulich (Eds.), *Principles and practice of aviation psychology* (pp. 115-147). Mahwah, NJ: Lawrence Erlbaum.
- Vitu, F., O'Regan, J.K., & Mittau, M. (1990). Optimal landing position in reading isolated words and continuous text. *Perception and Psychophysics*, 47, 583-600.
- VMBl (Ministerialblatt des Bundesministeriums der Verteidigung) (10.10.1990). Vorgaben zur Durchführung von Untersuchungen in der Bundeswehr (S. 403-404). Bundesministerium der Verteidigung (BMVg).
- Volpert, W. (1974). Handlungsstrukturanalyse als Beitrag zur Qualifikationsforschung. Pahl-Rugenstein.
- Watkins, M.J., Watkins, O.C., Craik, F.I.M., & Mazuryk, G. (1973). Effect of non-verbal distraction on short-term storage. *Journal of Experimental Psychology*, 101, 296-300.
- Weigel, H., Lindner, P., Henning, M., Kienast, H., Wanielik, G., Krems, J.F., & Spanner-Ulmer, B. (2009). A Novel Head Up Display for Driver Assistance Systems. Tagungsband 16th World Congress on Intelligent Transport Systems. In *Proceedings of ITS*. 21.- 25. September 2009, Stockholm, Schweden.

- Wetzel, P.A., Anderson, G.M., & Barelka, B.A. (1998). Instructor Use of Eye Position Based Feedback for Pilot Training. In Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting (pp. 1388-1392). Chicago, Illinois, USA.
- Wickens, C.D. (1984). Processing resources in attention. In R. Parasuraman & R. Davies (Eds.), *Varieties of attention* (pp. 63-101). New York: Academic Press.
- Wickens, C.D. (1986) The effects of control dynamics on performance. In K.R. Boff, L. Kaufman, & J.P. Thomas (Eds.), *Handbook of perception and performance Vol. II* (pp. 39–1/39–60). New York: John Wiley.
- Wickens, C.D. (1992). *Engineering Psychology and Human Performance* (2nd ed). New York: Harper-Collins.
- Wickens, C.D. (1996). Situation awareness: Impact of automation and display technology. In AGARD Conference Proceedings: *Situation Awareness: Limitations and Enhancement in the Aviation Environment* (pp. k2.1-k2.13), Brussels.
- Wickens, C.D. (2008). Multiple resources and mental workload. *Human Factors*, 50, 449-455.
- Wickens, C.D. (2009). Aviation psychology. Online-Dokument verfügbar unter: <http://e-book.lib.sjtu.edu.cn/iupsys/Proc/stock1/spv1ch24.html> [Zugriff: 09.04.2013].
- Wickens, C.D. & Hollands, J.C. (1999). *Engineering Psychology and Human Performance*. New York: Prentice Hall.
- Wickens, C.D., McCarley, J., & Thomas, L.C. (2004). Attention-Situation Awareness (A-SA) Model. In D. Folyle, B. Hooney, & A. Goodman (Eds.), *Human Performance Modeling Workshop Proceedings* (pp. 191-206). NASA Ames Research Center.
- Wierwille, W.A. (1993). Visual and manual demands of in-car controls and displays. In B. Peacock & W. Karwowski (Eds.), *Automotive Ergonomics* (pp. 299-320). Washington DC: Taylor & Francis.
- Wikman, A, Nieminen, T., & Summala, H. (1998). Driving experience and time-sharing during in-car tasks on road of different width. *Ergonomics*, 41(3). 358-372.
- Williams, K.W. (2000). *Impact of Aviation Highway-in-the-Sky Displays on Pilot Situation Awareness (Final Report)*. Washington DC, USA: Federal Aviation Administration (FAA).
- Williams, L.J. (1995). Peripheral target recognition and visual field narrowing in aviators and nonaviators. *International Journal of Aviation Psychology*, 5, 215-232.
- Wolpaw, J.R., Birbaumer, N., McFarland, D.J., Pfurtscheller, G., & Vaughan, T.M. (2002). Brain-computer interfaces for communication and control. *Clinical Neurophysiology*, 113, 767-791.
- Wouters, P., van Nimwegen, C., van Oostendorp, H., & van der Spek, E.D. (2013). A Meta-Analysis of the Cognitive and Motivational Effects of Serious Games. *Journal of Educational Psychology*. Advance online publication. DOI: 10.1037/a0031311.
- Xing, J. (2004). *Measures of information complexity and the implications for automation design*. Oklahoma, OK, USA: Civil Aerospace Medical Institute.
- Xing, J. & Manning, C A. (2005). *Complexity and automation displays of air traffic control: Literature review and analysis* (Rep. No. DOT/FAA/AM-05/4). Washington DC: FAA Office of Aerospace Medicine.
- Yerkes, R.M. & Dodson, J.D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, 18, 459-482.
- Zijlstra, F.R.H. & van Doorn, L. (1985). *The construction of a scale to measure subjective effort*. Technical Report, Delft University of Technology, Department of Philosophy and Social Sciences.
- Zwahlen, H.T., Adams, C.C., & DeBald, D.P. (1988). Safety aspects of CRT touch panel controls in automobiles. *Stress and Tension Control 3. Stress Management*, 335-344.

7 Anhang

7.1 Abkürzungsverzeichnis

α	Irrtumswahrscheinlichkeit
ACT-R	Adaptive Control Thought-Rational
ANOVA	Analysis of Variance
AOI	Area of Interest
AV	Abhängige Variable
BiV	Bildverstärker
BMVg	Bundesministerium der Verteidigung
Box M-Test	Box'scher M Test
bzw.	beziehungsweise
c	Cut-Off-Wert
C	Cargo
CH	Cargo Helicopter
cm	Zentimeter
df	Degrees of Freedom
DLR	Deutsches Zentrum für Luft- u. Raumfahrt e. V.
2D	Two-Dimensional
3D	Three-Dimensional
4D	Four-Dimensional
EASA	European Aviation Safety Agency
ebd.	ebenda
EC	Eurocopter
EDA	Elektrodermale Aktivität
EEG	Elektroenzephalogramm
e. g.	exempli gratia (zum Beispiel)
EMDR	Eye Movement Desensitization and Reprocessing
EOG	Elektrookulografie
f	Effektstärke (GPower)
F	Fighter
F	Fisher's F
FAA	Federal Aviation Administration
FAS	Fahrerassistenzsystem
fMRI	Functional Magnetic Resonance Imaging
FOV	Field of View
ft	Feet (Fuß; Flughöhe)
η	Effektstärke (Cohen, 1988)
GND	Ground
Gz-Kräfte	Beschleunigungskraft, die auf die dorsoventrale (z) Achse wirkt
HDD	Head Down Display
HGA	Hubschrauberführergrundausbildung
HMI	Human-Machine-Interaction
HMM	Hidden Markov Model
HR	Herzrate
HRV	Herzratenvariabilität
HUD	Head Up Display
IFR	Instrumental Flight Rules
ISO	Europäische Norm der International Standards Organisation
KIAS	Knots Indicated Airspeed
KS-Test	Kolmogorov-Smirnov-Anpassungstest
KS-Z	Standardisierte Prüfgröße des KS-Tests
LED	Licht emittierende Diode

M	Mittelwert
MANOVA	Multivariate Analysis of Variance
MCH	Modified Cooper-Harper Scale
min	Minuten
Mio.	Millionen
ML	Mechanically Limited
MMS	Mensch-Maschine-System
ms	Millisekunden
MW	Megawatt
n	Gruppengröße
N	Gesamtstichprobengröße
NASA-TLX	National Aeronautics and Space Administration Task-Load-Index
NATO	North Atlantic Treaty Organization
NL	Non-Localized
NM	Nautische Meile
OTL	Oberstleutnant
OTW	Out The Window
OWTD	Overwater Transition Down
p	Signifikanzniveau
PDF	Portable Document Format
PC	Personal Computer
PUI	Pupillenunruheindex
R	Bestimmtheitsmaß der multiplen Regression
r	Korrelation
RSME	Rating Scale Mental Effort
RTO	Research and Technology Organization
s.	siehe
S.	Seite
SA	Situation Awareness
SAGAT	Situation Awareness Global Assessment Technique
SAR	Safe and Rescue
SC	Situational Cognizance
SD	Standardabweichung
sec	Sekunden
SEEV-Model	Saliency-Effort-Expectancy-Value-Model
SL	Software Limited
SPSS	Statistical Package for the Social Sciences
SSQ	Simulator Sickness Questionnaire
SWAT	Subjective Workload Assessment Technique
SWORD	Subjective Workload Dominance Technique
UAV	Unmanned Aerial Vehicle
UH / UHT	Unterstützungshubschrauber / Tiger
US	United States
UV	Unabhängige Variable
VFR	Visual Flight Rules
vgl.	vergleiche
VMBI	Ministerialblatt
VOR	Vestibulo-Okular-Reflex
VP	Versuchsperson
vs.	versus
WP	Workload Profile
z. B.	zum Beispiel

7.2 Experteninterview

Experteninterview Bückeberg am 17.10.11 n=6

- Existieren innerhalb der HGA standardisierte Scanning Techniken? Wenn ja, welche?

Nein, keine vorgeschriebene
Flugführung anhand d. Displaywert / "Landwinkels"
bei Geländeflug / in Verbindung mit "Crosscheck"

- Werden die Scanning Techniken unter den Fluglehrern abgestimmt?

jeder hat seinen eigenen Stil

- Welche Scanning Techniken werden häufig falsch / abgewandelt durchgeführt?

Problem: Target Fixation; zu langes Fixieren im
Cockpit wird kompensiert, Parameter wird betrach-
tet; Sim. Sicht, ↓ Flugsicherheit

⇒ Lehrer schauen Display ab (Abdrehen von
Target Fixat.)

- Was sind typische schwierige bzw. komplexe Missions (z. B. hohe mentale
Workload = besonders viel rechnen; besonders viel mentales rotieren
erforderlich; viele Parallelaufgaben; viel Kommunikation; Arbeiten unter
Zeitdruck; Entscheidungen ohne ausreichend Information; viele visuelle
Checks notwendig; bei Geländeflug: Anzahl der Hindernisse usw.) innerhalb
der HGA?

- Geländeflug: Schwerkopf / Höhen; viel Spaly. witz,
Check Landwinkel, Landwinkels, Untergründ
- keine Referenzpunkte vorhanden im physischen Sichtfeld
- "Herausheben", Hindernisse, Höhe + IFR, Landwinkels im
Cockpit
- Was sind im Vergleich typische bzw. wenig komplexe Missions innerhalb der
HGA?
- viele Referenzpunkte, Landgründ was einsehbar
- Sichtumkehr (aber: "komplexe!")
- z. B. "Pinnacle"

- Wann ist eine Mission schwierig bzw. komplex?

keine Ref. Objekte, Herankommen, große Datenmenge / als Gedächtnis

- > schlechte Informationsaufnahme +
- Fehlende, neues Kopieren Kombinationen

- Welche Leistungsparameter bewerten Sie bei einem Flugschüler?

- Restkapazität
- Abwärtig. Landzeit (Kurs)
- Firmenschip
=> subjektiv!

- Welche Landeindizes / was genützt für Vordr?

1. Pinacole (Lücken)
2. Flugzeugträger (Nordsee)
3. Richtlinie (Damm)
4. Aufrang
5. conf. area (jetzt schief)

7.3 Rangliste

Ranking zum Eye Tracking Versuch (nur Fluglehrer!)

Bitte ordnen Sie als Experte die folgenden Landemanöver nach ihren visuellen Anforderungen. Vergeben Sie die Zahlen 1 bis 5. Jede Zahl sollte nur einmal zugeordnet werden. Dabei gilt:

- 1 = nicht schwierig
- 2 = wenig schwierig
- 3 = mittelmäßig
- 4 = eher schwierig
- 5 = sehr schwierig

Beachten Sie: Hohe visuelle Anforderungen bedeuten beispielsweise...

- ...es muss viel Information aus dem Cockpit aufgenommen werden, weil die Umgebung wenig Referenzen bietet.
- ...Information aus dem Cockpit muss schnell erfasst und verarbeitet werden, weil häufige Blickwechsel nach innen und außen erforderlich sind.
- ...das Einschätzen und Einhalten des Anflugwinkels ist schwer, weil die Umgebung nicht „fix“ oder das Gelände schwierig ist.

Landemanöver	Ranking
Pinnacle/ Kuppe (Loccum, 5 x 5 Meter, Geländeflug)	
Ridgeline/ Damm (Geländeflug)	
Gruppenversorger/ Fregatte (Nordsee)	
Absetzen/ Aufnehmen am Hang (Geländeflug)	
Confined Area (Geländeflug)	

Mit welchem Hubschrauber-Muster haben Sie die meisten Erfahrungen?

Antwort: _____

Vielen Dank!

© 2012 Dipl.-Psych. Maxi Robinski

7.4 Genehmigungsschreiben BMVg



Bundesministerium
der Verteidigung

Flugmedizinisches Institut der Luftwaffe
Abteilung IV - Ergonomie
Frau Dipl.-Psych. 'in Maxi Robinski
Flugplatz
85077 Manching

Niels Hanssen
P III 5

HALSANSCHRIFT Fontainengraben 150, 53123 Bonn
POSTANSCHRIFT Postfach 1328, 53003 Bonn
TEL +49(0)228-99-24-7472
FAX +49(0)228-99-24-1335
EMAIL Bmvgpii5@bmvg.bund.de

BEZUG: **Empirische Untersuchungen zur Einstellungs-, Meinungs- und Verhaltensforschung in der Bundeswehr**
hier: Genehmigung und Registrierung von Untersuchungsprojekten
ANLAGE - 1 - (VMBI)
Gez. P III 5 - Az 66-20-00
DATUM Bonn, 5. April 2012

Sehr geehrte Frau Robinski,

hiermit genehmige ich die von Ihnen beantragte Befragung.

Das Projekt wird unter der

Registriernummer 2/03/12

**Titel: „Eye Tracking und mentale Beanspruchung:
Die Rolle des Blickverhaltens im Cockpit“**

geführt.

Die Bestimmungen zur Durchführung von Untersuchungen in der Bundeswehr (VMBI vom 10.10.1990, Seiten 403 und 404) sind zu beachten (s. Anlage).



"VMBI 1990.pdf"

Aufgrund der Bestimmungen zum Datenschutz wird darauf hingewiesen, dass Befragungen nur auf freiwilliger Basis, Datenerhebungen, -speicherungen und -auswertungen ausschließlich anonym durchgeführt werden dürfen.

- 2 -

Ich bitte Sie sicherzustellen, dass vor der Durchführung der Untersuchung die Kommandeurin / der Kommandeur informiert wird und die Beteiligungsgremien vor Ort entsprechend einzubinden sind.

Nach Abschluss bitte ich um Vorlage eines Ergebnisberichts. In diesem Zusammenhang verweise ich besonders auf die Ziffer 8 des VMBI, nach der sich das BMVg vorbehält, einer Veröffentlichung der Ergebnisse nicht oder nur in Teilen zuzustimmen.

Mit freundlichen Grüßen
Im Auftrag
Hanssen

7.5 Probandeninformation

Registriernummer BMVg 2/03/12, P III 5 - Az 66-20-00
„Eye Tracking und mentale Beanspruchung: Die Rolle von Scanning Techniken im Cockpit“

Probandeninformation

Im Rahmen der Untersuchung weise ich Sie darauf hin, dass

- die Teilnahme an der Befragung freiwillig ist,
- Ihnen durch eine Verweigerung der Teilnahme keine Nachteile entstehen,
- die Erhebung und die Auswertung der Daten anonym, mindestens aber ohne offenkundigen Personenbezug, durchgeführt werden und die beabsichtigte Bekanntgabe des Untersuchungsergebnisses keine Einzeldaten enthält sowie keinen Rückschluss auf Einzelpersonen zulässt,
- die Erhebungsunterlagen - gegebenenfalls mit besonderem Sperrvermerk - nach Maßgabe des BArchG dem BA-MA zur Übernahme angeboten und, sofern das BA-MA auf die Unterlagen verzichtet, diese spätestens zwei Jahre nach Abschluss der Auswertung vernichtet werden.

Die Befragung findet im Rahmen eines Simulatorversuches statt, dessen Ablauf Ihnen bei einem Vortrag durch Dipl.-Psych´in Robinski erklärt wird. Die Daten aus der Befragung werden benötigt, um die Messergebnisse des Eye Tracking Systems anhand Ihrer subjektiven Bewertungen zu validieren. Ihre Angaben werden pseudonymisiert gespeichert und anonym ausgewertet. Nach erfolgter Auswertung wird ein Ergebnisbericht erstellt.

Ich danke vorab für Ihre Teilnahme.



Maxi Robinski

Maxi Robinski
Dipl.-Psych.
Flugmedizinisches Institut der Luftwaffe
Abteilung IV - Ergonomie
Flugplatz
85077 Manching
Tel. (08459) 80-3315
Fax (08459) 80-3400
Email: MaxiRobinski@bundeswehr.org

7.8 Fragebogen zur ergonomischen Bewertung der Eye Tracking Methode

PB Nr.	
Dienstgrad	
Datum	
Funktion	<input type="checkbox"/> Fluglehrer <input type="checkbox"/> Flugschüler
Brille?	<input type="checkbox"/> Nein <input type="checkbox"/> Ja
Kontaktlinsen?	<input type="checkbox"/> Nein <input type="checkbox"/> Ja
Akute Augenprobleme (z. B. Bindehautentzündung, verschwommene Sicht?)	<input type="checkbox"/> Nein <input type="checkbox"/> Ja Falls ja, welche?  _____
Flugstunden EC 135?	<input type="checkbox"/> 0 <input type="checkbox"/> 1-50 <input type="checkbox"/> 51-100 <input type="checkbox"/> 101-500 <input type="checkbox"/> 501-1000 <input type="checkbox"/> 1001-2000 <input type="checkbox"/> über 2000
Simulatorstunden EC 135?	<input type="checkbox"/> 0 <input type="checkbox"/> 1-50 <input type="checkbox"/> 51-100 <input type="checkbox"/> 101-500 <input type="checkbox"/> 501-1000 <input type="checkbox"/> 1001-2000 <input type="checkbox"/> über 2000

1. Ließ sich die Kamera-Brille komfortabel am Kopf **aufsetzen**?

Nein

 Warum nicht? _____

Ja

2. Ließ sich die Kamera-Brille komfortabel am Kopf **absetzen**?

Nein

 Warum nicht? _____

Ja

3. Empfanden Sie die **Fassung** der Kamera-Brille als **stabil**?

Nein

 Warum nicht? _____

Ja

4. War die Kamera-Brille im Bereich der **Nasenauflage** komfortabel?

Nein

 Warum nicht? _____

Ja

5. War die Kamera-Brille im Bereich der **Stirnauflage komfortabel**?
- Nein
 Warum nicht? _____
- Ja
6. War die Kamera-Brille im Bereich der **Auflage hinter den Ohren komfortabel**?
- Nein
 Warum nicht? _____
- Ja
7. **Verrutschte** die Kamera-Brille unter dem Helm, wenn Sie **den Kopf bewegt haben**?
- Nein
 Ja
 Bemerkungen? _____
8. Hatten Sie beim Tragen der Kamera-Brille **Missempfindungen** (z. B. Schmerzen, Drücken) unter dem Helm?
- Nein
 Ja
 Wo? _____
9. Hatten Sie beim Tragen eine **eingeschränkte Sicht nach außen** durch die Kamera-Brille?
- Nein
 Ja
 Nicht/ eingeschränkt gesehen: _____
10. Hatten Sie beim Tragen eine **eingeschränkte Sicht auf die Instrumente im Cockpit** durch die Kamera-Brille?
- Nein
 Ja
 Nicht/ eingeschränkt gesehen: _____
11. Hat Sie die Kamera-Brille beim Tragen an einer **sicheren Flugführung gehindert**?
- Nein
 Ja
 Warum? _____
12. Wie beurteilen Sie den **Komfort** beim Tragen der Kamera-Brille unter dem Helm **zusammenfassend**?
- optimal
 - fast optimal
 - akzeptabel
 - akzeptabel mit Einschränkungen
 - nicht akzeptabel

13. Wie beurteilen Sie den **Diskomfort** beim Tragen der Kamera-Brille unter dem Helm **zusammenfassend** (Erklärung: körperliches oder psychisches Unwohlsein, Schmerzen, Müdigkeit, Unruhe, Beanspruchung)?

- optimal
- fast optimal
- akzeptabel
- akzeptabel mit Einschränkungen
- nicht akzeptabel

14. Spüren Sie im Nachhinein **Druckstellen** durch die Kamera-Brille? Wenn ja, wo?

- Nein
- Ja

 Wo? _____

15. Spüren Sie im Nachhinein **Juckreiz** durch die Kamera-Brille?

- Nein
- Ja

 Wo? _____

16. Spüren Sie im Nachhinein **Schmerzen** durch die Kamera-Brille?

- Nein
- Ja

 Wo? _____

17. Hat Sie die **Verkabelung** mit dem Eye Tracking System bei der sicheren Flugführung gestört?

- Nein
- Ja

 Warum? _____

18. Hat Sie die **Anwesenheit der Versuchsleiterin** im Simulator bei einer sicheren Flugführung gestört?

- Nein
- Ja

 Warum? _____

19. Waren Symptome von Simulator Sickness **durch das Eye Tracking System** verstärkt ausgeprägt?

- Nein, nicht stärker als sonst.
- Ja, stärker als sonst.

 Welche (bitte ankreuzen)?⁴⁵

⁴⁵ Items aus Simulator Sickness Questionnaire (SSQ; Kennedy, Lane, Berbaum, & Lilienthal, 1993).

	gar nicht	gering	mittel	schwer
Allgemeines Unwohlsein				
Ermüdung				
Kopfschmerzen				
Überanstrengte Augen				
Schwierigkeiten mit Sehschärfe				
Erhöhter Speichelfluss				
Schwitzen				
Übelkeit				
Konzentrationsschwierigkeiten				
Druckgefühl im Kopfbereich				
Verschwommene Sicht				
Schwindel (bei geöffn. Augen)				
Schwindel (bei geschl. Augen)				
Gleichgewichtsstörungen				
Magenbeschwerden				
Aufstoßen				

21. Wie **häufig** könnten Techniken wie Eye Tracking **als Feedback-Methode** eingesetzt werden?

- Vor/nach jedem Simulatorflug
- Regelmäßig einmal pro Woche
- Regelmäßig einmal pro Monat
- Einmal in der gesamten HGA
- Gar nicht

 Warum nicht? _____  Sonstige Anm.: _____

20. Abgesehen vom technischen Stand des Eye Tracking Systems (Verkabelung); eignet sich Eye Tracking als **Feedback-Methode** beim **Realflug**?

- Nein

 Warum nicht? _____

- Ja

Vielen Dank für Ihre Teilnahme!

© 2012 Dipl.-Psych. Maxi Robinski

7.9 Validitätsindizes für Pupille und Marker

Tabelle 22: Validitätsindizes für Pupillen- und Markervalidität (N = 33).

Proband Nr.	Mission 1 (Pinnacle)			Mission 2 (Gruppenversorger)		
	Eye Validity [%]	Marker Validity [%]	Dauer der Mission [min]	Eye Validity [%]	Marker Validity [%]	Dauer der Mission [min]
1	92.94	93.89	5:57	93.92	42.23	3:53
2	97.31	75.63	5:51	97.34	78.97	3:22
3	99.06	95.55	4:24	99.43	97.46	4:48
4	98.84	71.77	4:52	97.98	68.48	4:33
5	92.78	90.16	4:30	95.62	88.42	3:09
6	97.14	40.38	1:50	97.85	14.45	3:41
7	90.40	82.30	4:04	95.96	80.32	2:21
8	98.87	2.41	4:21	97.61	36.07	2:42
9	97.77	80.96	3:39	97.75	85.73	4:22
10	96.53	51.82	4:32	96.04	20.10	4:58
11	92.90	63.08	3:49	95.49	64.85	3:33
12	98.78	83.04	5:11	98.35	84.68	4:12
13	98.96	92.54	4:14	90.03	77.12	2:49
14	98.60	70.07	5:55	98.96	77.75	3:31
15	98.48	71.43	4:38	96.71	62.61	4:21
16	97.84	76.07	4:34	97.95	82.14	4:19
17	98.24	96.45	3:39	98.06	79.04	3:30
18	99.44	63.05	4:22	99.17	92.84	3:36
19	99.10	90.94	3:54	96.97	78.40	3:20
20	99.11	79.15	4:21	99.26	78.23	4:24
21	99.20	90.32	3:43	99.03	62.99	2:33
22	95.73	75.23	3:55	97.17	78.10	4:25
23	97.89	95.65	4:37	98.96	64.10	4:14
24	99.04	78.33	5:28	99.14	91.41	4:51
25	96.27	84.51	4:00	95.16	93.93	3:31
26	98.07	73.62	3:58	96.90	83.12	4:30
27	95.52	82.17	3:30	92.91	82.25	3:59
28	96.97	64.16	4:05	95.19	61.65	4:18
29	96.88	84.82	4:17	99.22	61.60	3:09
30	99.05	96.73	3:14	97.29	80.72	4:35
31	97.04	95.33	4:14	95.60	66.72	3:58
32	92.29	88.50	4:34	97.05	50.63	4:12
33	96.52	50.50	6:14	96.59	55.47	5:13

Anmerkung: Die **markierten** Datensätze unterschritten das Markervaliditätskriterium von 70 % und wurden von der Datenanalyse ausgeschlossen.

7.10 Ergebnisse zum NASA-TLX

Tabelle 23: Subskalen des NASA-TLX (N = 66).

Subskala	Mentale Anforderungen	Physische Anforderungen	Zeitliche Anforderungen	Performance	Anstrengung	Frustration
M	52.88	20.08	6.36	83.26	47.88	36.52
SD	24.92	15.85	15.48	15.45	28.50	35.26

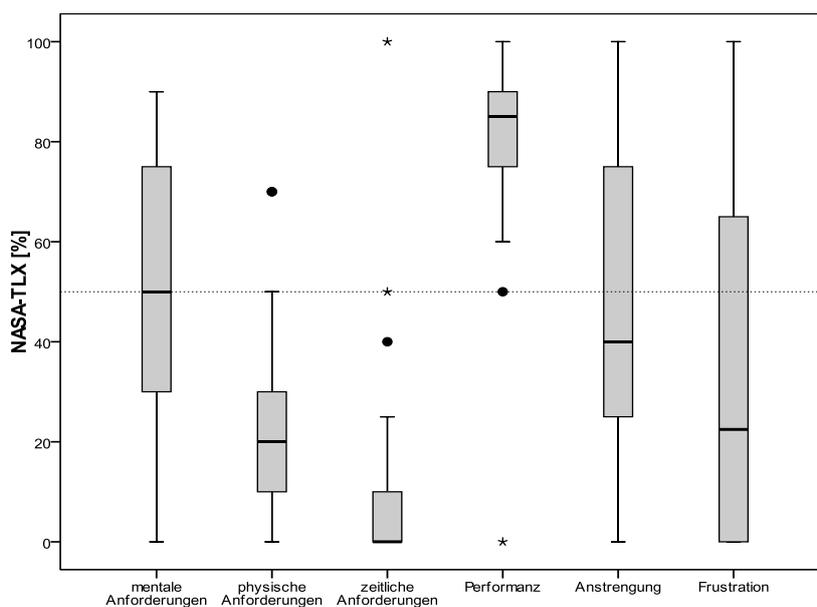


Abbildung 62: Subskalen des NASA-TLX (N = 66).

Tabelle 24: T-Test für die Subskalen des NASA-TLX auf dem Faktor Erfahrung.

Subskala	Levene-Test	F	Sig.	T	Df	p
Mentale Anforderungen	Varianzen sind gleich	0.00	.998	-2.55	64.00	.013
	Varianzen sind nicht gleich			-2.55	63.64	.013
Physische Anforderungen	Varianzen sind gleich	9.94	.002	1.37	64.00	.176
	Varianzen sind nicht gleich			1.39	51.86	.170
Zeitliche Anforderungen	Varianzen sind gleich	6.59	.013	1,50	64.00	.137
	Varianzen sind nicht gleich			1,54	40.94	.131
Performance	Varianzen sind gleich	0.21	.646	-0.17	64.00	.865
	Varianzen sind nicht gleich			-0.17	59.91	.864
Anstrengung	Varianzen sind gleich	0.41	.525	-1.69	64.00	.096
	Varianzen sind nicht gleich			-1.69	63.92	.096
Frustration	Varianzen sind gleich	0.76	.386	-0.05	64.00	.964
	Varianzen sind nicht gleich			-0.05	62.61	.964

Anmerkung: T-Test für unabhängige Stichproben mit α -Adjustierung nach Bonferroni; n_{Lehrer} = 34, n_{Schüler} = 32.

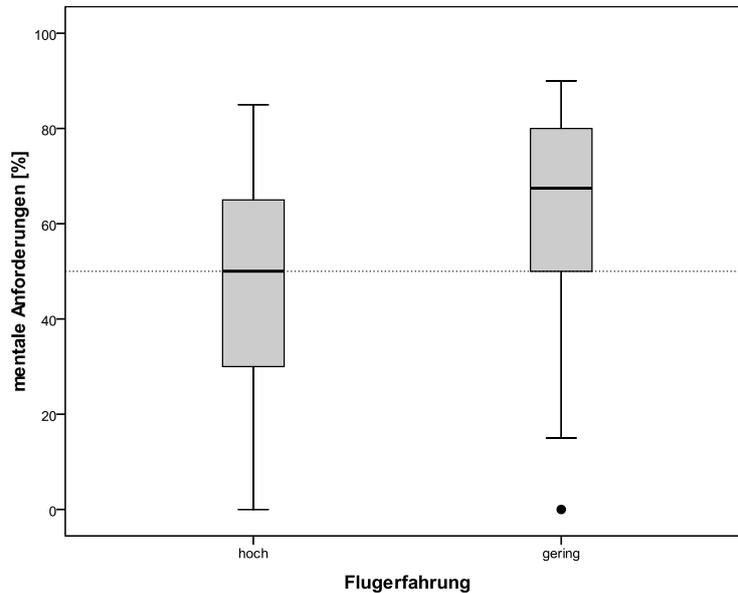


Abbildung 63: Unterschied für die Subskala mentale Anforderungen auf dem Faktor Erfahrung.

7.11 Post Hoc Analysen zur MANOVA

Dargestellt sind die Follow Up ANOVAs (α -Adjustierung nach Bonferroni) zum signifikanten Interaktionseffekt der MANOVA für die Variable Target Fixations zur Detektion von signifikanten Einzelunterschieden in den Mittelwerten. Die signifikante Differenz liegt in der Gruppe der Fluglehrer auf dem Faktor visuelle Anforderungen (s. Tabelle 25, Tabelle 26, Abbildung 64). Alle weiteren Tests erbrachten keine signifikanten Unterschiede (s. Tabelle 27, Tabelle 28).

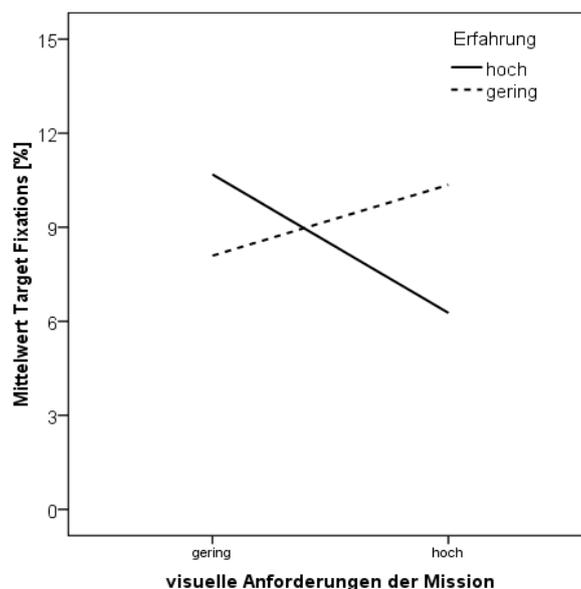


Abbildung 64: Target Fixations auf dem Faktor visuelle Anforderungen.

Tabelle 25: Target Fixations auf dem Faktor visuelle Anforderungen.

Subgruppe	Visuelle Anforderungen	n	M	SD
Fluglehrer	Gering	16	10.68	5.13
	Hoch	9	6.26	4.32
	Gesamt	25	9.09	5.23
Flugschüler	Gering	10	8.10	3.27
	Hoch	10	10.35	6.33
	Gesamt	20	9.22	5.04

Tabelle 26: Follow Up ANOVA auf dem Faktor visuelle Anforderungen.

Subgruppe		Quadratsumme	df	Mittel der Quadrate	F	p
Fluglehrer	Between	112.29	1	112.29	4.75	.040
	Within	543,86	23	23.65		
	Total	656.15	24			
Flugschüler	Between	25.52	1	25.52	1.01	.329
	Within	457.01	18	25.39		
	Total	482.53	19			

Tabelle 27: Target Fixations auf dem Faktor Erfahrung.

Subgruppe	Erfahrung	n	M	SD
Hohe visuelle Anforderungen	Hoch	9	6.26	4.32
	Gering	10	10.35	6.33
	Gesamt	19	8.41	5.72
Geringe visuelle Anforderungen	Hoch	16	10.68	5.13
	Gering	10	8.10	3.27
	Gesamt	26	9.69	4.61

Tabelle 28: Follow Up ANOVA auf dem Faktor Erfahrung.

Subgruppe		Quadratsumme	df	Mittel der Quadrate	F	p
Hohe visuelle Anforderungen	Between	79.09	1	79.09	2.63	.123
	Within	510.12	17	30.00		
	Total	589.21	18			
Geringe visuelle Anforderungen	Between	41.22	1	41.22	2.02	.169
	Within	490.75	24	20.45		
	Total	531.97	25			

8 Lebenslauf der Autorin

Der Lebenslauf wurde aus datenschutzrechtlichen Gründen entfernt.